AWARD NUMBER: W81XWH-16-1-0130

TITLE: Single-Cell Dissection of Human Pancreatic Islet Dysfunction in Diabetes

PRINCIPAL INVESTIGATOR: Michael L. Stitzel, Ph.D.

RECIPIENT: The Jackson Laboratory
Bar Harbor, ME 04609

REPORT DATE: June 2017

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
June 2017	Annual	1 Jun 2016 - 31 May 2017

4. TITLE AND SUBTITLE

Single-Cell Dissection of Human Pancreatic Islet Dysfunction in Diabetes

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-16-1-0130

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Michael L. Stitzel, Ph.D.

E-Mail: Michael.Stitzel@jax.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The Jackson Laboratory for
Genomic Medicine
10 Discovery Drive
Bar Harbor, ME 04609

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Islets are composed of ≥5 endocrine cell types that perform complementary functions to maintain proper glucose homeostasis. This cellular heterogeneity impedes our ability to understand the precise transcriptional repertoire and regulatory landscape of each cell type and to determine how these programs in each cell type are perturbed in type 2 diabetes (T2D). The overarching goal of this project is to determine, with single cell resolution, changes in cellular composition and cell-specific gene expression programs elicited by T2D in human islets using innovative single cell transcriptomic (scRNA-seq; Aim 1) and epigenomic (scATAC-seq; Aim 2) technologies.

15. SUBJECT TERMS
Single cell; epigenome; scATAC-seq; scRNA-seq; transcriptome; human islet; alpha; beta; delta; pancreatic polypeptide (PP); gamma; epsilon; endocrine

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	55	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

1. INTRODUCTION:

Islets are composed of ≥5 endocrine cell types that perform complementary functions to maintain proper glucose homeostasis. This cellular heterogeneity impedes our ability to understand the precise transcriptional repertoire and regulatory landscape of each cell type and to determine how these programs in each cell type are perturbed in type 2 diabetes (T2D). The overarching goal of this project is to determine, with single cell resolution, changes in cellular composition and cell-specific gene expression programs elicited by T2D in human islets using innovative single cell transcriptomic (scRNA-seq; Aim 1) and epigenomic (scATAC-seq; Aim 2) technologies.

2. KEYWORDS:

Single cell; epigenome; scATAC-seq; scRNA-seq; transcriptome; human islet; alpha; beta; delta; pancreatic polypeptide (PP); gamma; epsilon; endocrine

3. ACCOMPLISHMENTS:

What were the major goals of the project?

Major goals of the project:
Aim 1: Islet single cell transcriptomes
1a: Non-diabetic (ND) islet single cell transcriptomes
Milestone (12 months): ~1000 single cell transcriptome profiles from 5 ND islets
Achieved: 4806 single cell transcriptome profiles from 1 ND islet (~500% of cells, 20% of samples to date)

1b: Type 2 diabetic (T2D) islet single cell transcriptomes
Milestone (12 months): ~1000 single cell transcriptomes from 5 T2D islets
Achieved: ~4192 single cell transcriptome profiles from 1 T2D islet (~400% of cells, 20% of samples to date)

1c: Determine cell type transcriptome signatures in ND and T2D samples
Milestone (anticipated, month 15): Comprehensive analysis of islet transcriptomes and identification of cell type-specific transcriptomes / "signature" genes
Achieved: Identification of cell type-specific transcriptomes and signature genes from two samples

1d: Identify cell type-specific expression differences between ND and T2D samples
Milestone (anticipated, month 15): Identification of cell type-specific differential expression in T2D vs. ND samples

Aim 2: Islet single cell epigenomes
2a: Non-diabetic (ND) islet single cell epigenomes
Milestone (12 months): ~1000 single cell epigenome (scATAC-seq) profiles
Achieved: ~400 single cell epigenome profiles (40%)

2b: Type 2 diabetic (T2D) islet single cell epigenomes
Milestone (12 months): ~1000 single cell epigenome (scATAC-seq) profiles
Achieved: ~400 single cell epigenome profiles (40%)

2c: Determine cell type epigenome signatures in ND and T2D samples
Milestone (anticipated, 15 months): Comprehensive analysis of islet epigenomes and identification of cell type-specific regulatory element use/epigenome signatures

Achieved to date: Established and evaluated scATAC-seq analysis pipelines; QC and unsupervised clustering analyses of 1 ND islet

2d: Identify cell type-specific epigenomic differences between ND and T2D samples
Milestone (anticipated. 15 months): Identification of cell type-specific differences in regulatory element use/epigenome signatures in T2D and ND states
Achieved to date: Identification of aggregate single cell ATAC-seq regulatory element signatures from 1 ND islet

What was accomplished under these goals?

1) *Major activities*: We have completed single cell transcriptome and epigenome profiling of 1 T2D and 1 matched ND islet to date. We have effectively implemented analysis pipelines for both islet single cell transcriptome and epigenome analyses.

2) *Specific objectives*: Specific objectives in this reporting period were (1) to establish and test analytical pipelines for single cell transcriptome and epigenome analysis and (2) complete single cell profiling of pancreatic islets from 5 non-diabetic and 5 type 2 diabetic (T2D) donors.

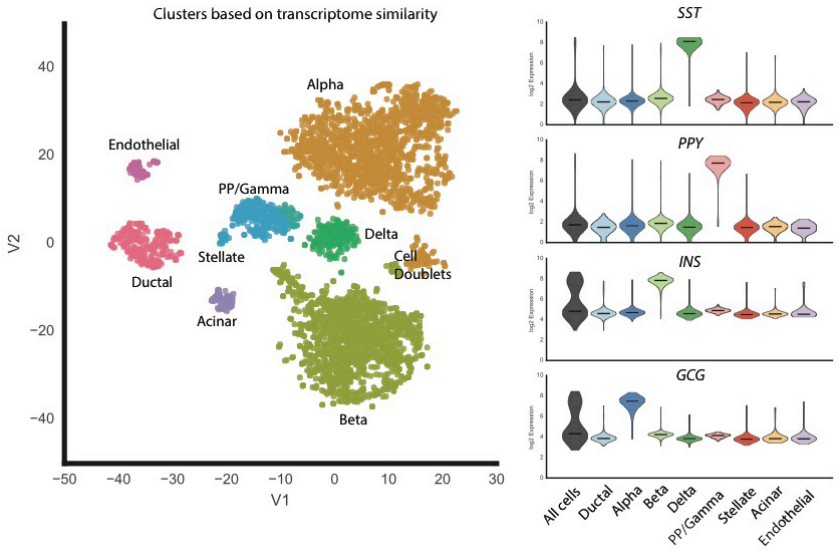


Figure 1. Single cell transcriptomes of ND islet cluster into major cell types with little evidence for subpopulations. (Left): t-SNE dimensional reduction analysis of single cell transcriptome similarity. Cells clustered into major endocrine and exocrine cell types as labeled. (Right): Violin plots of marker gene expression for delta (*SST)*, gamma (*PPY*), beta (*INS*), and alpha (*GCG)* cells. Note: color codes in right panel do not correlate to those for cell clusters on the left.

3) *Significant results:* We have completed single cell transcriptome profiling of 4806 cells from one ND islet and 4192 cells from one T2D islet using the 10X Genomics platform. As shown in **Figures 1 and 2**, we have identified all major endocrine cell populations in both ND and T2D samples. Importantly, the ability to profile thousands of single cell transcriptomes per donor for the same cost as the hundreds of cells per donor originally proposed is allowing us to determine more rigorously (1) if there are subpopulations of endocrine cells in ND And T2D islets, (2) if new, rarer populations of cells such as "de-differentiated beta cells" are appearing in T2D samples, and (3) if the relative known and novel cell (sub)populations are changing in proportion between ND and T2D islets. Analyses of our single cell transcriptome data are limited to n=1 each for ND and T2D states, so we refrain thus far from making broad conclusions. However, our results so far suggest the following:

- Endocrine cells in ND islets cluster by cell type (alpha, beta, delta, gamma/PP). We neither find evidence of endocrine cell type subpopulations nor of novel populations in this ND islet (**Fig. 1**).
- T2D islets contain discernible alpha and beta cell subpopulations **(Fig. 2A,** Beta 1,2 and Alphas 1-3**)**. We are still in the process of analyzing these data, but our preliminary differential analyses of Beta 1 and Beta 2 subpopulations (**Fig. 2A-D**) indicate that Beta 2 expresses similar levels of *INS* as Beta 1 (**Fig. 2B**). However, Beta 2 exhibit significantly lower *MAFA* (**Fig. 2C**), *NKX6-1* (**Fig. 2D**), and *PDX1*, *MAFB*, and *NEUROD1* (not shown) expression, consistent with previous targeted observations by Roland Stein's group[1]. Moreover, we detect significant induction of genes modulating stress responses such as amyloid responses, processing, and degradation (*APLP2, APP, ITMB2)* inflammation (*CTSA, RBP4)*, autophagy *(TMEM59, LAMP1, LAMP2, ATP6AP2*), and oxidative and endoplasmic reticulum stress responses (*PRDX4, ERP29*) in this same population. Finally, we identified differential regulation

of T2D GWAS effector genes *ZFAND6*, *CDKN2A*, and *KCNK17* between these two beta cell populations. We are continuing these analyses through the next reporting period and will extend them to identify the genes and pathways distinguishing the three alpha cell subpopulations. Most importantly, we will complete these analyses in additional islet samples to identify robust and reproducible T2D cell type-specific signatures among multiple individuals.

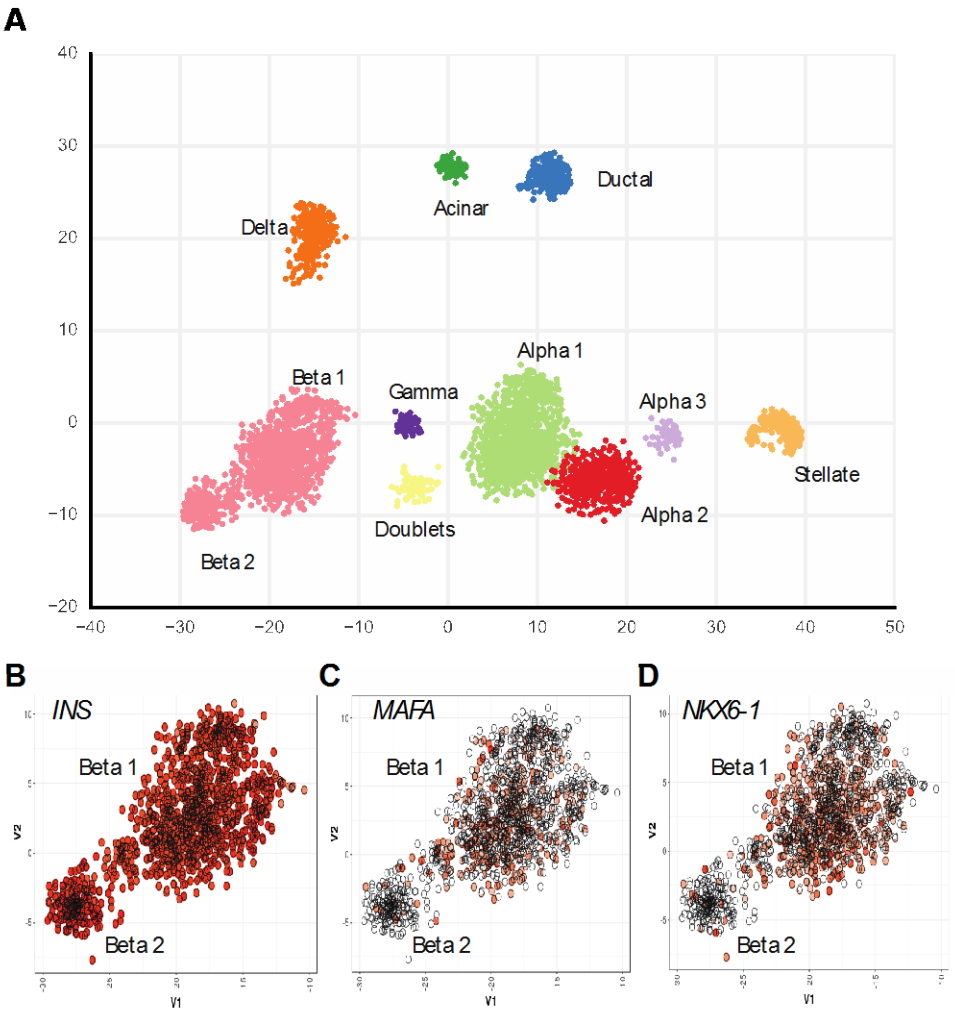


Figure 2. Single cell transcriptomes reveal putative alpha and beta cell subpopulations in T2D islets. (A) t-SNE analysis of 4192 single cell transcriptomes from one T2D islet donor. **(B-D)** Beta 2 subpopulation exhibits similar high levels of *INS* expression (**B, pink/red dots**) but significantly lower *MAFA* (**C**) and *NKX6-1* (**D**) expression than Beta 1.

Epigenome profiling of single cells using ATAC-seq to identify open chromatin sites has been completed for one T2D and one ND individual. As shown in **Fig. 3**, aggregate scATAC-seq profiles from 165 single cells identify many of the consensus sites identified in bulk islet samples ("Islet ATAC-seq") (gray boxes). Importantly, these include islet-specific sites (compare GM12878 ATAC-seq to "Islet ATAC-seq" and "Aggregate scATAC-seq" plots) at both transcription start sites and distal regulatory elements, suggesting our profiling approach is capturing the full range of regulatory element classes (e.g., promoter, enhancer, insulator) in islets. The sparse nature of the scATAC-seq profiles in each specific single cell has made assigning each single cell epigenome profile to a specific cell type (e.g., alpha, beta, delta, PP/gamma) challenging. We believe part of this difficulty is due to technical issues of over- or under-transposition of samples. To address this technical challenge, we are completing serial dilution experiments to determine optimal transposase concentrations that will yield robust scATAC-seq profiles for

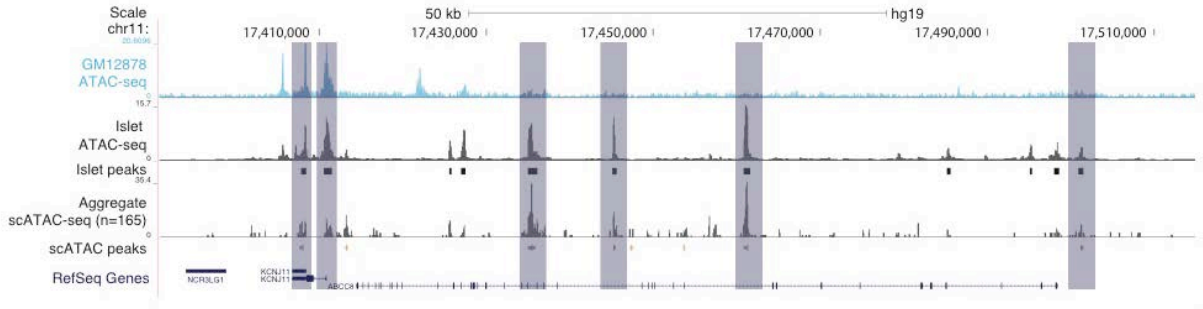


Figure 3. Aggregation of single cell ATAC-seq profiles largely recapitulates the epigenomic landscape identified in bulk islet samples. UCSC Genome Browser view of the *KCNJ11/ABCC8* genes encoding the sulfonylurea receptor subunits. Gray boxes highlight open chromatin sites consistently identified by bulk islet and aggregated single cell (n=165 cells) ATAC-seq profiles. Several of these sites are islet-specific and not observed in a lymphoblastoid cell line (GM12878). Gene models (RefSeq Genes) are shown at the bottom.

each single cell. We do not anticipate that this is an insurmountable technical hurdle. Pending the results of these experiments, however, we may explore an alternative approach to transpose multiple pools of tens to hundreds of FACS-enriched alpha, beta, delta, and gamma cells each to generate more robust cell type-specific epigenome profiles from ND and T2D islet donors. This alternative approach would only be implemented after consultation with and approval by Dr. Thakar, Program Officer on this Discovery Award, as it would represent a change in the scope of the project, namely changing the epigenome profiling assay resolution of from single cell to pools of a specific enriched cell type.

Stated goals not met: In this reporting period, we have experienced challenges in obtaining the number of islets originally proposed. This was due to less frequent availability of T2D islets (5 offers total between August 2016 and May 2017) than anticipated based on frequency in the previous two years of tracking (~1 offer per month). Moreover, for 3 T2D offers, the islet preparation at ProdoLabs or IIDP failed to yield high quality islets. To address this challenge in obtaining T2D islets, we discussed and implemented an alternative islet subscription strategy with IIDP to maximize our chances to obtain T2D islets. Since its implementation in May, we have received 3 targeted offers, for which one donor met our inclusion criteria. We have identified and profiled an appropriately matched non-diabetic donor. As described in *Major activities* and *Significant results* above, our experimental and analytical pipelines are robust and generating high quality profiles, so we are equipped and prepared to complete expedited analyses of the islet datasets in months 13-18 . Given the current rate of obtaining islets, we anticipate we can obtain islets from 3-4 T2D donors from the same number of matched ND donors and analyze them within the upcoming 5-6 months. If we experience another lag in T2D islet availability in the upcoming 1-2 months, we will seek alternative sources of islets, including the University of Alberta center directed by Dr. Patrick MacDonald.

What opportunities for training and professional development has the project provided?

Nothing to Report

How were the results disseminated to communities of interest?

Nothing to Report

What do you plan to do during the next reporting period to accomplish the goals?

To deal with the lag in obtaining T2D islets due to lower than anticipated availability of T2D islets and isolation failures at the IIDP/ProdoLab centers, (noted in 3.3 above), we have specifically discussed and implemented a strategy with the IIDP coordinating team to maximize targeted offers for T2D islets and match non-diabetic donor islets using the Open offers. These cells are still acquired under the same protocols used by the IIDP/ProdoLabs and approved by OHRP. After implementing this change, we have obtained a well-matched T2D-ND set of donors in the span of 1 month. Moreover, our adaptation of the 10X Genomics platform for single cell RNA-sequencing allows us to sample ~4000-6000 single cell transcriptomes per sample, which will maximize our ability to detect cellular subpopulations and determine quantitative differences between these cell types/subpopulations within a given sample (as shown in **Fig. 2**) and between T2D and ND samples.

1c: Determine cell type transcriptome signatures in ND and T2D samples
Milestone (anticipated, month 15): Comprehensive analysis of islet transcriptomes and identification of cell type-specific transcriptomes / "signature" genes
Plans moving forward: We believe that cell type-specific transcriptomes will not differ substantially between ND donors, and therefore anticipate that we have already identified the cell type-specific transcriptomes of healthy islets. However, based on our initial T2D islet results, we could conceivably identify different

subpopulations among different T2D donors. With the updated islet procurement strategy, we anticipate to be able to obtain and sequence the remaining four T2D donor islets by month 15-16. Analysis pipelines are in place and robust, so we expect that analyzing the additional data should take no longer than 2-3 weeks, which should keep us within our original plans to summarize and write up study results by month 18.

1d: Identify cell type-specific expression differences between ND and T2D samples
Milestone (anticipated, month 15): Identification of cell type-specific differential expression in T2D vs. ND samples
Plans moving forward: As noted above, we anticipate obtaining and profiling all needed T2D and ND samples by months 16 and 17, respectively. Differential analyses between specific cell types (e.g. T2D delta vs. ND delta) should take approximately 1-2 weeks additional time.

2c: Determine cell type epigenome signatures in ND and T2D samples
Milestone (anticipated, 15 months): Comprehensive analysis of islet epigenomes and identification of cell type-specific regulatory element use/epigenome signatures
Plans moving forward: As noted in Significant Results, aggregate scATAC-seq profiles reflect those of bulk islets, but the sparse nature of these epigenomic datasets have made assigning each single cell ATAC-seq profile to a specific cell type challenging. We have transposed two additional samples and will determine if changing transposase concentration results in more robust scATAC-seq profiles in the next month. If it does not, we will contact Dr. Thakar and discuss an alternative approach, namely ATAC-seq of each endocrine cell type enriched by fluorescence-activated cell sorting (FACS) to define alpha, beta, delta, and PP/gamma cell type epigenomes in ND and T2D islets. As these samples will be processed in parallel with the transcriptome profiling (*Aim 1d)*, we anticipate the same timeline to reach completion.

2d: Identify cell type-specific epigenomic differences between ND and T2D samples
Milestone (anticipated. 15 months): Identification of cell type-specific differences in regulatory element use/epigenome signatures in T2D and ND states
Plans moving forward: The specific strategy to meet a timeline of completed data collection and analyses of ~16-17 months is detailed for *Aim 2c*, above. We have completed differential analyses of whole islet ATAC-seq profiles from ND and T2D donors, thus the pipeline is in place and analysis should ≤ 2 weeks longer than the time needed to profile and sequence the samples.

4. IMPACT: Describe distinctive contributions, major accomplishments, innovations, successes, or any change in practice or behavior that has come about as a result of the project relative to:

What was the impact on the development of the principal discipline(s) of the project?

Nothing to Report

What was the impact on other disciplines?

Our single cell studies and datasets led to some collaborative discussions with Dr. Zhijin Wu at Brown University. She is developing novel approaches to analyze single cell data to reveal new and unique insights into the dynamics and mechanisms of gene expression at the single cell level. The results of this collaboration have yielded co-authorship on a manuscript describing their methodology, which is in preparation for submission to *Bioinformatics*.

What was the impact on technology transfer?

Nothing to Report

What was the impact on society beyond science and technology?

Data from this study were shared in the Community Health Discussion series, a community outreach initiative involving The Jackson Laboratory and The Children's Museum in West Hartford. In this educational presentation for the general public, I presented initial results from our single cell islet transcriptome analyses and explained how they can reshape our understanding of pathogenic events/processes contributing to diabetes and how they may shift our approach to preventing and treating diabetes.

5. CHANGES/PROBLEMS:

Changes in approach and reasons for change

Nothing to Report

Actual or anticipated problems or delays and actions or plans to resolve them

As noted above, we encountered unforeseen delays obtaining T2D islets due to lower than usual availability and unproductive islet yields in ProdoLabs/IIDP islet distribution centers. In consultation with the IIDP coordinating center, we have implemented a new strategy to prioritize T2D islets in targeted offers and to obtain matched non-diabetic islet offers via the open offers. In the span of one month, this strategy yielded islets from a well-matched pair of T2D and ND donors. Thus, we are confident this plan will result in higher rates of procurement over the next 3-5 months that should allow us to meet originally proposed milestones and ultimate deliverables of cell type-specific differential expression (Aim 1d) and cell type-specific differences in regulatory element use (Aim 2d) in T2D vs, ND samples central to this proposal.

Changes that had a significant impact on expenditures

We have incurred less material and experimental costs than budgeted in this current reporting period due to the delays in obtaining suitable islets matching our inclusion criteria. As we increase our T2D islet procurement rate (and the corresponding matched ND samples), we anticipate that these decreased costs will normalize to our projected budget in the upcoming months.

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents:

Significant changes in use or care of human subjects

Nothing to Report

Significant changes in use or care of vertebrate animals.

Not applicable; project does not involve the use of vertebrate animals.

Significant changes in use of biohazards and/or select agents

Nothing to Report

6. PRODUCTS: List any products resulting from the project during the reporting period. If there is nothing to report under a particular item, state "Nothing to Report."

- **Publications, conference papers, and presentations**
  Report only the major publication(s) resulting from the work under this award.

  Journal publications. *List peer-reviewed articles or papers appearing in scientific, technical, or professional journals. Identify for each publication: Author(s); title; journal; volume: year; page numbers; status of publication (published; accepted, awaiting publication; submitted, under review; other); acknowledgement of federal support (yes/no).*

1. Lawlor N, Khetan S, Ucar D, and **Stitzel ML**. Genomics of Islet (Dys)function and Type 2 Diabetes. Trends Genet. 2017 Apr;33(4):244-255. PMID28245910;
   Acknowledgement of federal support: Yes
   Please see Appendix

2. Lawlor N, George J, Bolisetty M, Kursawe R, Sun L, V S, Kycia I, Robson P, **Stitzel ML**. Single cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. 2016. Genome Res. Nov 18. [Epub ahead of print] PMID: 27864352
   Acknowledgment of federal support: Yes
   Please see Appendix

3. Wu Z, Zhang Y, Stitzel ML, and Wu H. Two-phase differential expression analysis for single cell RNA-seq. *Under review, Bioinformatics.*

4. Bolisetty M, Stitzel ML, and Robson, P. CellView: Interactive Exploration of High Dimensional Single Cell RNA-seq Data. *Under review, Bioinformatics.*
   BioRxiv link: (http://biorxiv.org/content/early/2017/04/04/123810)
   Please see Appendix

Books or other non-periodical, one-time publications.

Nothing to Report

Other publications, conference papers, and presentations.

Nothing to Report

- **Website(s) or other Internet site(s)**

Nothing to Report

- **Technologies or techniques**

Nothing to Report

- **Inventions, patent applications, and/or licenses**

Nothing to Report

- **Other Products**

Nothing to Report

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name:	*Michael Stitzel*
Project Role:	*PD/PI*
Research Identifier:	*http://orcid.org/0000-0001-5630-559X*
Nearest person month worked:	*1*
Contribution to Project:	*Dr. Stitzel has managed the project, directing both the experiments and analyses completed by Drs. Kursawe and Mr. Lawlor*

Name:	*Joshy George*
Project Role:	*Co-Investigator/Computational Scientist*
Researcher Identifier:	*http://orcid.org/0000-0001-8510-8229*
Nearest person month worked:	*1*
Contribution to Project:	*Dr. George has set up pipelines and trained Mr. Lawlor to complete the computational analyses.*

Name:	*Romy Kursawe*
Project Role:	*Research Assistant IV*
Research Identifier:	
Nearest person month worked:	*4*
Contribution to Project:	*Dr. Kursawe has completed all of the experiments for the project, including processing islets, preparing single cell suspensions, preparing RNA, transposing nuclei, and preparing ATAC-seq libraries*

Name:	*Nathan Lawlor*
Project Role:	*Data Analyst*
Research Identifier:	*http://orcid.org/0000-0003-3263-6057*
Nearest person month worked:	*1*
Contribution to Project:	*Mr. Lawlor has implemented and runs the published analysis pipelines for single cell ATAC-seq and single cell RNA-seq and internal pipelines established by Dr. George. In addition, he has completed differential analyses between cell (sub)populations and has participated in analyses and content for peer-reviewed publications supported by this funding*

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Yes. Changes in Other Support for Key Personnel are italicized below.

Stitzel, Michael L

Active

Supporting Agency:	NIH/NIDDK 5 R00 DK092251-05	**PI:**	Stitzel
Project Title:	Investigation of noncoding variation in human pancreatic islets and their developmental precursors		
Role:	Principal Investigator	**Effort:**	6.00 CM
Entire Project:	08/20/2014 - 07/31/2017	$794,250	
Current Year:	08/01/2016 - 07/31/2017	$249,000	
Project Goals:	The goal of this research project is to understand the role that genetic variation in non-protein coding regulatory regions of the genome play in human pancreatic islet function and dysfunction.		
Specific Aims:	1. Determine which of approximately 40 candidate regulatory regions in the human genome function as enhancers, silencers, or insulators in human islets; 2. Determine which elements contain variants that alter gene expression in adult human islets; 3. Determine which elements contain variants that alter gene expression in pancreatic precursor cells.		
Overlap:	None		
Contracting/ Grants Officer:	Sheryl Sato - satos@extra.niddk.nih.gov		

Supporting Agency:	*Department of Defense* *W81XWH-16-1-0130*	***PI:***	*Stitzel*
Project Title:	*Single-Cell Dissection of Human Pancreatic Islet Dysfunction in Diabetes*		
Role:	*Principal Investigator*	***Effort:***	*0.50 CM*
Entire Project:	*06/01/2016 - 11/30/2017*	*$350,000*	
Current Year:	*06/01/2016 - 11/30/2017*	*$350,000*	
Project Goals:	*The goal of this project is to test the hypothesis that non-diabetic (ND) and T2D islets exhibit distinct cell type-specific transcriptomic and/or epigenomic signatures that are masked by the cellular heterogeneity in whole islet studies.*		
Specific Aims:	*Aim 1: Identify cell type-specific transcriptome signature differences between non-diabetic (ND) and T2D* *islets using single cell RNA-sequencing (scRNA-seq); Aim 2: Identify cell type-specific epigenome (open chromatin) signature differences between ND and T2D islets using single cell assay for transposase-accessible chromatin using sequencing (scATAC-seq).*		
Overlap:	*None*		
Contracting/ Grants Officer:	*Lisa Sawyer - lisa.m.sawyer22.civ@mail.mil*		

Completed

The Jackson Laboratory Director's Innovation Fund JAX-DIF-FY15-DUJB
"Advancing ATAC-seq Data Generation and Analysis Pipeline for Epigenetic Biomarker Discovery"
PI: Banchereau/Ucar
Role: Co-Investigator

The Jackson Laboratory Director's Innovation Fund TJL-DIF-FY14-GRHGWC
"Maximizing Human and Mouse Resources to Identify Novel Variants for Alzheimer's Disease"
PI: Carter / Howell
Role: Co-Investigator

The Jackson Laboratory Director's Innovation Fund TJL-DIF-FY14-GWC
"Genetics of Molecular Epigenetics"
PI: Carter
Role: Co-Investigator

George, Joshy

Active

Supporting Agency:	NIH/NCI 5 R01 CA195712-03	**PI:**	Flavell / Palucka
Project Title:	Humanized mouse models to dissect in vivo the interplay between melanoma and the immune system		
Role:	Co-Investigator	**Effort:**	0.60 CM
Entire Project:	05/14/2015 - 04/30/2018	$1,304,550	
Current Year:	05/01/2017 - 04/30/2018	$395,352	
Project Goals:	The goal of this project is to credential the humanized MISTRG mouse model as a platform for investigating the immune-mediated mechanisms of tumorigenesis by establishing transcriptional signatures linked with melanoma progression and confirming these signatures in tumors from patients.		
Specific Aims:	1. Determine the architecture of human melanoma tumors and their impact on human tumor-infiltrating immune cells in vivo in MISTRG mice reconstituted		

	with donor CD34+ HPCs and melanoma cell lines; 2. Define how human melanoma alters the human systemic immunity in MISTRG mice; 3. Validate the MISTRG model in an autologous system where MISTRG mice are reconstituted with patient CD34+ HPCs and autologous tumors.
Overlap:	None
Contracting/ Grants Officer:	Debra Sowell - debra.sowell@nih.gov

Supporting Agency:	*NIH/NIA 5 R01 AG052608-02*	*PI:*	*Banchereau*
Project Title:	*Genomics and Epigenomics of the Elderly Response to Pneumococcal Vaccines*		
Role:	*Co-Investigator*	***Effort:***	*0.96 CM*
Entire Project:	*09/01/2016 - 04/30/2020*	*$2,348,313*	
Current Year:	*05/01/2017 - 04/30/2018*	*$546,881*	
Project Goals:	*The goal of this project is to dissect the age-related changes in immune cells that affect responses to microbial vaccination in the elderly.*		
Specific Aims:	*1. To vaccinate healthy elderly with two distinct pneumococcal vaccines, collect longitudinal blood samples and assess pneumococcal-specific antibody responses; 2. To establish the transcriptional and epigenetic profiles of elderly blood immune cells linked with antibody responses to pneumococcal vaccination; 3. To examine the functional status of immune cells in the elderly stratified according to their pneumococcal vaccine responder status.*		
Overlap:	*None*		
Contracting/ Grants Officer:	*Mitchell Whitfield - whitfieldm@od.nih.gov*		

Supporting Agency:	*NIH/NIAID 5 R01 AI121920-02*	***PI:***	*Unutmaz*
Project Title:	*Decoding Immunological perturbations during Chronic Fatigue Syndrome*		
Role:	*Co-Investigator*	***Effort:***	*1.80 CM*
Entire Project:	*06/01/2016 - 05/31/2021*	*$3,281,517*	
Current Year:	*06/01/2017 - 05/31/2018*	*$642,090*	
Project Goals:	*The goal of this project is to develop a detailed functional and genetic immunological framework that can be used to decode the mechanisms of Myalgic Encephalomyelitis and Chronic Fatigue Syndrome (ME/CFS) and to develop robust, quantitative immune-biomarker sets for predicting disease susceptibility, stratifying patients and guiding treatment strategies.*		
Specific Aims:	*1) To determine the frequencies of immune cell subsets in the blood of a clinically defined ME/CFS patient cohort; 2) To assess functional capacity of memory T cells, innate cells and the differentiation potential of naive T cells during ME/CFS; and 3) To determine the T cell and innate cell subset?specific gene and lncRNA transcripts in ME/CFS patient blood samples.*		
Overlap:	*None*		
Contracting/ Grants Officer:	*LeBrit Nickerson - lebrit.nickerson@nih.gov*		

Supporting Agency:	*NIH/NIAID 5 U01 AI124297-02*	***PI:***	*Banchereau*
Project Title:	*Combination Adjuvants to Activate Human Dendritic Cell Subsets and B Cells*		
Role:	*Co-Investigator*	***Effort:***	*0.96 CM*

Entire Project:	*04/15/2016 - 03/31/2021*	*$3,177,983*
Current Year:	*04/01/2017 - 03/31/2018*	*$640,154*
Project Goals:	*Our goal is to select a combination adjuvant using functional assays, followed by in-depth investigation of molecular pathways accounting for enhanced immunogenicity. Our collaboration with industry will enable the transition of the selected combination adjuvant to further studies of human vaccination.*	
Specific Aims:	*1. Select a combination adjuvant that induces maximal humoral immunity in vitro; 2. Identify the molecular mechanisms through which the selected combination adjuvant activates DCs; 3. Examine the adjuvant effect in vivo and validate the selected pathway using CRISPR-CAS9 engineering of human CD34+HPCs.*	
Overlap:	*None*	
Contracting/ Grants Officer:	*LeBrit Nickerson - lebrit.nickerson@nih.gov*	

Supporting Agency:	*Department of Defense W81XWH-16-1-0130*	***PI:***	*Stitzel*
Project Title:	*Single-Cell Dissection of Human Pancreatic Islet Dysfunction in Diabetes*		
Role:	*Co-Investigator*	***Effort:***	*0.40 CM*
Entire Project:	*06/01/2016 - 11/30/2017*	*$350,000*	
Current Year:	*06/01/2016 - 11/30/2017*	*$350,000*	
Project Goals:	*The goal of this project is to test the hypothesis that non-diabetic (ND) and T2D islets exhibit distinct cell type-specific transcriptomic and/or epigenomic signatures that are masked by the cellular heterogeneity in whole islet studies.*		
Specific Aims:	*Aim 1: Identify cell type-specific transcriptome signature differences between non-diabetic (ND) and T2D islets using single cell RNA-sequencing (scRNA-seq); Aim 2: Identify cell type-specific epigenome (open chromatin) signature differences between ND and T2D islets using single cell assay for transposase-accessible chromatin using sequencing (scATAC-seq).*		
Overlap:	*None*		
Contracting/ Grants Officer:	*Lisa Sawyer - lisa.m.sawyer22.civ@mail.mil*		

Supporting Agency:	*Department of Defense W81XWH-17-1-0010*	***PI:***	*Palucka*
Project Title:	*Molecular Mechanisms of Human TNBC Metastasis In Vivo*		
Role:	*Computational Scientist*	***Effort:***	*0.60 CM*
Entire Project:	*02/01/2017 - 01/31/2020*	*$1,386,446*	
Current Year:	*02/01/2017 - 01/31/2018*	*$460,079*	
Project Goals:	*Our overall goal is to test the hypothesis that TNBC metastatic dissemination is driven by a specific cancer-immune cell interaction that can be identified and targeted for therapy.*		
Specific Aims:	*1. Establish the capacity of TNBC PDXs to disseminate and form metastasis in hNSG-SGM3 mice; 2. Identify immune cells and their molecular signatures linked with TNBC PDX metastasis.*		
Overlap:	*None*		
Contracting/ Grants Officer:	*Susan Dodd - susan.l.dodd4.civ@mail.mil*		

Completed

None

What other organizations were involved as partners?

Nothing to Report

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS: For collaborative awards, independent reports are required from BOTH the Initiating PI and the Collaborating/Partnering PI. A duplicative report is acceptable; however, tasks shall be clearly marked with the responsible PI and research site. A report shall be submitted to **https://ers.amedd.army.mil** for each unique award.

QUAD CHARTS: If applicable, the Quad Chart (available on https://www.usamraa.army.mil) should be updated and submitted with attachments.

9. APPENDICES:

a) Lawlor N, Khetan S, Ucar D, and **Stitzel ML**. Genomics of Islet (Dys)function and Type 2 Diabetes. Trends Genet. 2017 Apr;33(4):244-255. PMID28245910;

b) Lawlor N, George J, Bolisetty M, Kursawe R, Sun L, V S, Kycia I, Robson P, **Stitzel ML**. Single cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. 2016. Genome Res. Nov 18. [Epub ahead of print] PMID: 27864352

c) Bolisetty M, Stitzel ML, and Robson, P. CellView: Interactive Exploration of High Dimensional Single Cell RNA-seq Data. *Under review, Bioinformatics.*

CellPress

Review

Genomics of Islet (Dys) function and Type 2 Diabetes

Nathan Lawlor,[1] Shubham Khetan,[1,2] Duygu Ucar,[1,3] and Michael L. Stitzel[1,2,3,*]

Pancreatic islet dysfunction and beta cell failure are hallmarks of type 2 diabetes mellitus (T2DM) pathogenesis. In this review, we discuss how genome-wide association studies (GWASs) and recent developments in islet (epi)genome and transcriptome profiling (particularly single cell analyses) are providing novel insights into the genetic, environmental, and cellular contributions to islet (dys) function and T2DM pathogenesis. Moving forward, study designs that interrogate and model genetic variation [e.g., allelic profiling and (epi)genome editing] will be critical to dissect the molecular genetics of T2DM pathogenesis, to build next-generation cellular and animal models, and to develop precision medicine approaches to detect, treat, and prevent islet (dys)function and T2DM.

Lay of the Land: (Functional) Genomic Landscape of Islets and T2DM

T2DM is a complex metabolic disorder with both genetic and environmental components. It results from the dysfunction and loss of insulin-secreting beta cells in the endocrine pancreas (Islets of Langerhans) as they work to secrete more insulin to counteract insulin resistance in peripheral tissues (adipose, skeletal muscle, and liver). Ultimately, T2DM manifests as uncontrolled elevations in blood glucose levels. **GWAS** (see Glossary) have systematically identified hundreds of **single nucleotide variants** (SNVs), representing >150 regions of the genome (loci) [1], that are associated with T2DM risk and differences in T2DM-related quantitative metabolic traits, such as insulin, proinsulin, and glucose levels. Most (>90%) of these SNVs reside in noncoding regions of the genome. In parallel, functional (epi)genomics approaches to map open chromatin using **DNase I hypersensitive site sequencing** (DNase-seq), **assay for transposase-accessible chromatin sequencing** (ATAC-seq), and histone modification and transcription factor (TF)-binding patterns using **chromatin immunoprecipitation sequencing** (ChIP-seq) have identified genome-wide location of regulatory elements (REs), such as promoters, enhancers, and insulators, in >150 human cell types and tissues. T2DM SNVs are significantly and specifically enriched in islet-specific REs [2–7], suggesting that changes in islet RE activity and target gene expression are a common mechanism underlying the molecular genetics of islet dysfunction and T2DM [8] (Figure 1A). Indeed, recent studies have identified putative factors binding these REs and have detected allelic effects on their binding and target gene expression [9–11].

In this review, we discuss how recent studies are improving our understanding of how islet REs are perturbed by SNVs contributing to T2DM risk [1,12–19] and are elucidating the transcriptional underpinnings of islet responses to (patho)physiological environmental changes, such as aging, circadian rhythms, Western diet and lifestyle, as well as oxidative, endoplasmic reticulum (ER), and inflammatory stress responses [20–25]. We explore how studies applying next-generation sequencing (NGS) to profile individual cells are improving our comprehension of islet biology and reshaping our view of T2DM pathogenesis. Finally, we examine similarities and differences between mice and humans in the 'omics of islet function and T2DM (summarized in

Trends

T2DM is a multi-tissue metabolic disorder that results when pancreatic islets fail to compensate for insulin resistance in peripheral tissues.

Recent studies reaffirm the common variant origins of T2DM genetic risk. Variants overlap noncoding genomic regions, implicating regulatory defects in T2DM etiology.

Environmental stressors are associated with changes in gene expression programs leading to T2DM progression.

Single cell sequencing technologies permit investigation of islet cell type transcriptomes and epigenomes with single cell resolution and/or precision. Such methods provide greater insight into cell type-specific perturbations and their roles in T2DM.

Recent studies suggest that other cells (alpha, delta, and PP/gamma) in the islet have important roles in islet/and/or beta cell function, resilience, and T2DM pathogenesis.

[1]The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032, USA
[2]Department of Genetics & Genome Sciences, University of Connecticut, Farmington, CT 06032, USA
[3]Institute for Systems Genomics, University of Connecticut, Farmington, CT 06032, USA

*Correspondence:
michael.stitzel@jax.org (M.L. Stitzel).

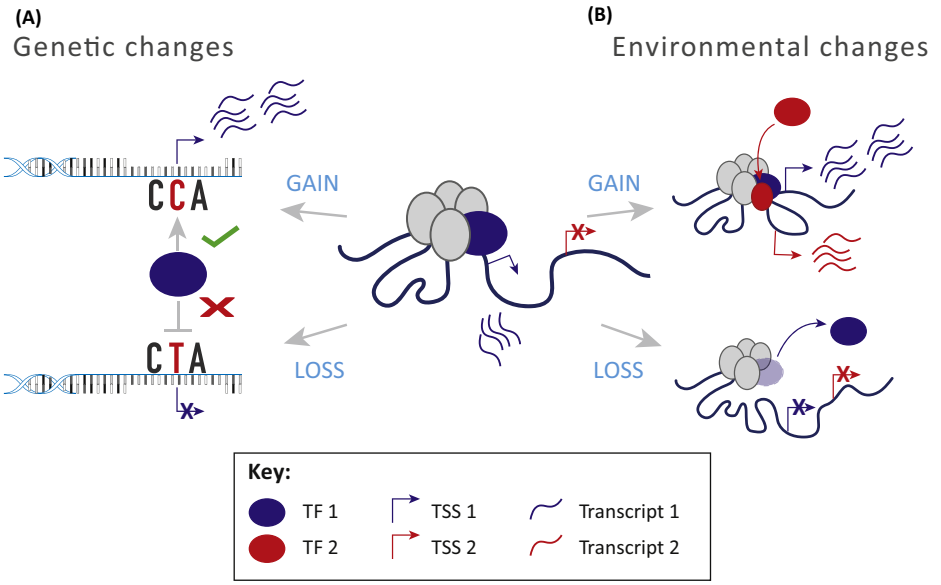


Figure 1. Genomic Effects of Genetic and Environmental Perturbations Contributing to Pancreatic Islet Dysfunction and Type 2 Diabetes Mellitus (T2DM). (A) DNA single nucleotide variants (SNVs) may enhance (gain-of-function) or diminish (loss-of-function) transcription element (e.g., enhancer) activity and islet gene expression. Most T2DM-associated SNVs reside in noncoding regions of the genome and overlap islet regulatory elements (REs) [2,3,12,14,15,32,47], implicating disruptions in gene regulatory network components as a central molecular feature in disease pathogenesis. A subset of SNVs has been linked to changes in basal islet gene expression [11,31]. (B) Environmental factors, such as inflammation, diet, aging, circadian rhythms, and stress, may also influence RE activity, resulting in altered and/or novel transcription of genes essential for islet function [20–25,48–50,57,58]. Abbreviations: TF, transcription factor; TSS, transcription start site

Figure 2, Key Figure). Throughout, we highlight future challenges and opportunities and offer perspectives on how these recent developments set the stage for precision medicine approaches to understand, treat, and prevent T2DM.

Homing in on T2DM Genetic Risk and Architecture

Since initial T2DM GWAS reports in 2007 [26–29], the list of genomic loci in which sequence variation contributes to T2DM risk and variability in quantitative measures of pancreatic islet function has grown to over 150 [1,14,30]. Associated SNVs at each locus contribute modestly to increased T2DM risk [odds ratios (OR) 1.05–1.75]. Together, these loci only explain a fraction of T2DM heritability [13,14]. Genetic consortia continue to dissect the genetic architecture of T2DM using larger cohorts with increasing ethnic diversity and/or representation. Recent efforts have reported [12,14,30,99] fewer 'new' T2DM loci (N=10) than previous studies. Importantly, however, they are refining the genetic signals at known (previously associated) T2DM loci to define **'credible sets' of single nucleotide polymorphisms** (SNPs) that are the most probable causal and/or functional SNPs driving the association and, consequently, the resulting molecular and/or phenotypic consequences.

The GOT2D and T2D-GENES consortia sought to identify less common SNVs (0.1%<MAF<5%) with larger effect size that may underlie common variant associations or may account for some of the T2DM 'missing heritability' using a combined whole-genome sequencing (WGS), exome sequencing, and genotype imputation approach [14]. These efforts identified protein-coding variants and/or mutations that are the most likely causative variant or effector transcripts for 12 out of 78 GWAS loci, confirming five nominated in previous studies (*PPARG, KCNJ11-ABCC8, SLC30A8, GCKR,* and *PAM* loci) and identifying seven

Glossary

Assay for transposase-accessible chromatin sequencing (ATAC-seq): a technique used to profile regions of open chromatin from small cell numbers.
Chromatin immunoprecipitation sequencing (ChIP-seq): a method used to study DNA–protein interactions.
Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET): a method used to study 3D chromatin interactions genome wide.
CpG sites: areas of DNA containing a cytosine nucleotide directly linked to a single phosphate group and guanine nucleotide. These sites are often methylated and influence transcription.
Credible sets of SNPs: lists of sequence variants with 95% posterior probability of containing a/the causal disease-associated SNP (s) [99].
Deconvolution: a statistical framework to resolve a heterogeneous mixture into its constituent elements.
Dedifferentiation: the process in which a mature differentiated cell type reverts to an earlier developmental and/or precursor state.
DNA methylation: molecular process wherein a methyl group is covalently attached to a DNA base without altering the DNA sequence.
DNase I hypersensitive site sequencing (DNase-seq): a method used to characterize regulatory and open chromatin regions of the genome.
Expression quantitative trait loci (eQTL): approach to link sequence variation at a position in the genome to expression of target gene(s).
Genome-wide association study (GWAS): statistical association of sequence variation with disease risk or variability in a measurable phenotypic trait and/or feature.
Glycated hemoglobin (HbA1C): a type of hemoglobin modification that is measured to determine plasma glucose concentration.
RNA-sequencing (RNA-seq): measures the amount of RNA in a sample at a given time.
Single nucleotide polymorphism (SNP): nucleotide variation at a specific location in the genome that exists with >5% frequency in the population.

new ones (*FES, TM6SF2,* and *RREB1* in the *PRC1, CILP2,* and *SSR1* loci, respectively, and *TSPAN8, THADA, HNF1A,* and *HNF4A*). For the remaining loci, noncoding SNVs constitute the putative causal SNVs. Comparison of multiple genetic models with the empirical data generated in this study suggest that a long tail of common variants with lower effect sizes may comprise the missing heritability and reaffirms the importance of common, regulatory variation in the genetic architecture of T2DM (see Outstanding Questions). Perhaps most importantly, this immense effort has narrowed the list of putative causal SNVs to a handful for five loci and by 50% on average for the 78 T2DM-associated autosomal loci investigated [14]. Similar themes and reductions in credible sets were reported for fasting glucose- and insulin-associated loci [30].

Ongoing islet epigenomic and transcriptomic analyses are progressively defining the regulatory potential of variant loci, identifying SNV-RE overlaps, and nominating potential target genes, whose dysfunction is likely to contribute to T2DM [2,3,11,12,14,15,30–32]. Open chromatin (DNase-seq, ATAC-seq) and histone modification and/or TF-binding profiling (ChIP-seq) indicate that T2DM and related trait-associated SNVs are especially prominent in islet distal REs and **stretch/super enhancers** [2,3,5,33,34]. Due to the long distances over which REs might act, additional work to elucidate the target genes of T2DM SNV-containing REs is needed. Chromosome conformation capture techniques, such as 3C, 4C, 5C [35], Hi-C [36], **chromatin interaction analysis by paired-end tag sequencing** (ChIA-PET) [37], and HiCHIP [38] will be important components to effectively map interactions between REs and their target genes (see Outstanding Questions). In two separate studies, **RNA-sequencing** (RNA-seq) of 89 [31] and 118 [11] human islet samples identified 616 and 2341 **expression quantitative trait loci** (eQTLs), respectively. These analyses were the first studies linking SNVs to gene expression changes in islets to define the putative genetic control of islet function and failure. However, of the 216 eQTLs common to both studies, only 14 overlapped with T2DM-associated loci [11]. This may be due to power limitations and an inability to detect eQTLs beyond their primary signal. Alternatively, this relatively low overlap could suggest that T2DM SNVs affect islet physiological or pathophysiological responses, not just basal expression, as has been measured to date. Indeed, a recent study suggested that several putative T2DM GWAS genes are regulated by NFAT, a TF involved in calcineurin signaling responses [39]. Alternatively, the detection of eQTLs overlapping T2DM-associated SNVs in peripheral tissues, such as skeletal muscle [40] and adipose [41] tissue, reminds us that these other metabolic tissues should not be ignored in the T2DM molecular genetics and pathogenesis, and warrant further investigation of genomic variation in these tissues.

Recent islet studies suggest that regulatory noncoding RNAs (ncRNAs) contribute to diabetes progression and beta cell (dys)function [31,42,43]. Aberrant expression of 17 long noncoding (lncRNAs) has been associated with **glycated hemoglobin** (HbA1c) levels [31]. This study identified eQTLs for two of these transcripts (*LOC283177* and *SNHG5*), but the eQTL SNVs did not overlap with T2DM SNVs [31]. Similarly, a study by Morán and colleagues identified nine out of 55 T2DM-associated loci that contained lncRNAs located within 150 kb of, but not directly overlapping, the reported lead SNVs [42]. In the *KCNQ1* locus, T2DM risk SNVs overlap both *KCNQ1* and *KCNQ1OT1* [43,44], a long intergenic noncoding RNA (lincRNA) also found to be significantly induced in T2DM islets [42]. We anticipate that additional links will emerge in the coming years. Other studies suggest that islet lncRNA alterations could also contribute to type 1 diabetes mellitus (T1DM), because a T1DM GWAS SNV (rs941576) was identified in the *MEG3* lincRNA locus [43,45]. Functional analyses in human islets and rodent models will clarify the roles of these ncRNAs in islet development, (dys)function, and diabetes.

DNA methylation studies of nondiabetic (ND) and T2DM islets have suggested that epigenetic dysregulation promotes T2DM development [46,47]. DNA methylation profiling of 15 T2DM and

Single nucleotide variant (SNV): changes in a given nucleotide sequence in the genome.
Stretch/super enhancers: extended (>3 kb) regions of the genome marked by enhancer chromatin states; enriched near genes that are important for cell type identity and cell type-specific functions.
Subpopulation: a subset of cells within a tissue distinguished by the expression of specific marker genes and/or proteins.
***Trans*-differentiation:** the process in which a mature cell type converts into another mature cell type.

Key Figure

Converging and Diverging Genetic, Environmental, and Cellular Aspects of Islet (Dys)function and Type 2 Diabetes Mellitus (T2DM) in Mice and Humans

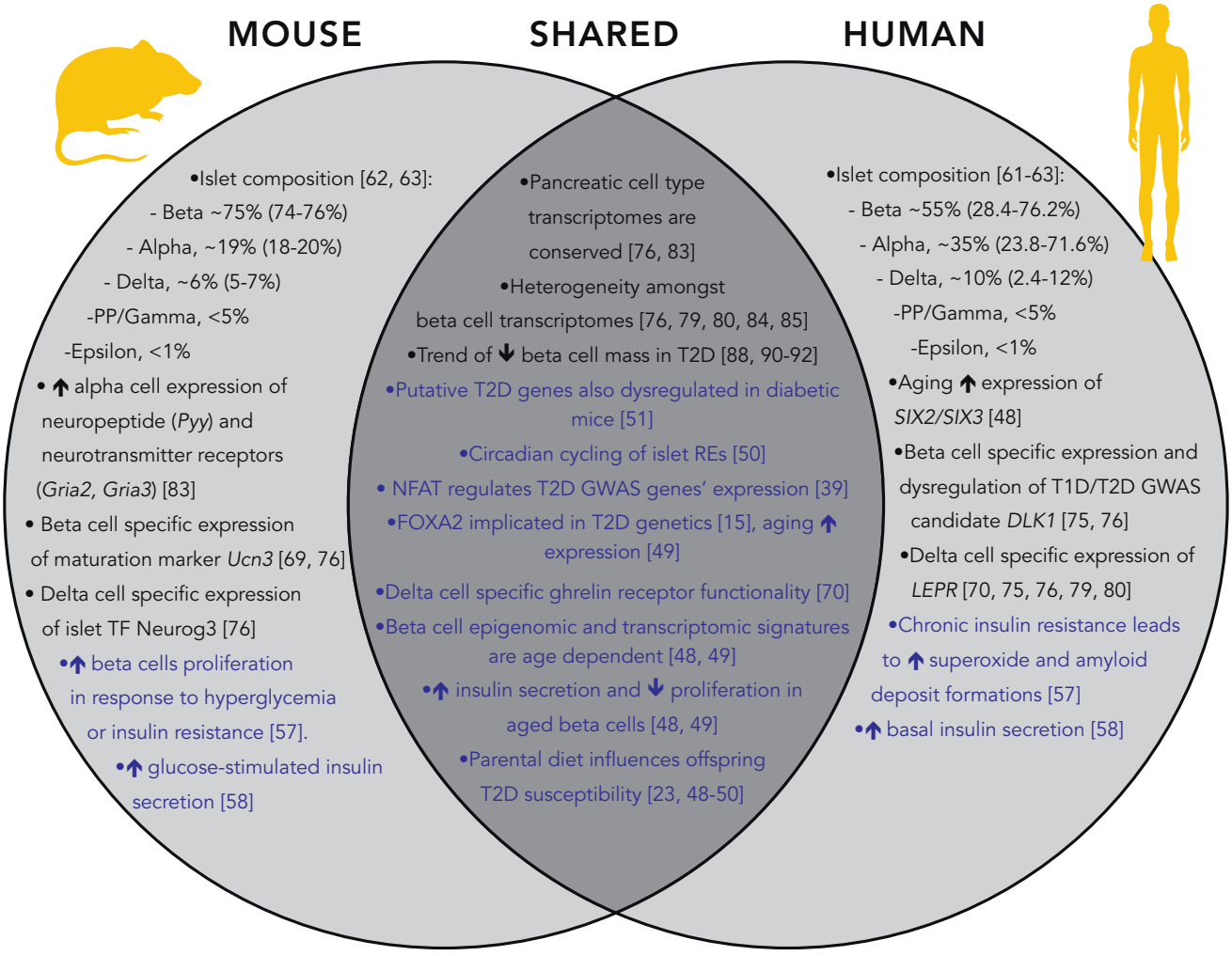


Figure 2. Parallel analyses of human and mouse islets are revealing important similarities [15,23,39,48–51,70,76,79,80,83–85] (A) and differences [48,57–58,61–63,69-70,75-76,79–80,83] (B,C) between molecular features of islet identity and (dys)function in mice and humans. Black text highlights significant findings regarding islet cellular composition and identity. Blue text highlights longitudinal and/or comparative analyses of genome-wide molecular data sets and environmental effects on islet (dys)function. These features reaffirm the value of modeling T2DM in mice to delineate important species-specific differences in islet biology that may reflect distinct T2DM causative mechanisms. Abbreviations: ↑, increase; ↓, decrease; GWAS, genome-wide association study RE, regulatory element; TF, transcription factor; T1DM, type 1 diabetes mellitus

34 ND islets using the Illumina 450BeadChip identified 1649 differentially methylated **CpG sites** for 853 genes, 17 of which reside in T2DM-associated loci [46]. Surprisingly, most (97%) of these CpG sites were hypomethylated in T2DM islets, suggesting that they suffer from decreased methyl donor levels or decreased activity of DNA methyltransferases.

Genomics of Islet Responses to Environmental Changes and T2DM Pathogenesis

Intrinsic and extrinsic environmental changes, such as aging, and Western diet and/or lifestyle, respectively, are linked to islet dysfunction and T2DM risk [23,48–50] (Figure 1B). Multiple groups have begun to characterize the genomic effects of these environmental inputs and insults on islets. Transcriptome profiling of adult and juvenile islet beta cells identified 565 (209 up, 356 down) and 6123 (2083 up, 4040 down) differentially expressed genes in humans and mice, respectively [48,49]. Signatures of decreased proliferative capacity in aged islets and/or beta cells were apparent in both species, perhaps best illustrated by increased *CDKN2A/B* expression, a gene cluster with established cellular senescence functions and implicated as 'Type 2 Diabetogenes' for a T2DM GWAS signal on 9p21 [48,49,51]. Unexpectedly, transcriptome and epigenome signatures suggested superior insulin secretory capacity of adult islets, which was confirmed functionally by glucose-stimulated insulin secretion (GSIS) assays [48,49]. DNA methylation and histone profiling indicated that these expression differences were largely mediated by chromatin remodeling and epigenetic modification of distal Res, such as enhancers. Using whole-genome bisulfite sequencing (WGBS), Avrahami and colleagues identified approximately 14 368 aging-related differentially methylated regions (DMRs) between the beta cells of juvenile and adult mice. DMRs overlapping distal REs outnumbered those overlapping promoters 3:1, and exhibited larger changes in magnitude of methylation. Distal DMRs that lost methylation with aging were enriched for binding sites of important islet TFs, such as Foxa2, Neurod1, and Pdx1, suggesting these factors mediate the expression differences and improved functionality in adult islets. Finally, genes showing differential expression in adult islets were accompanied by differential methylation at nearby distal REs more often than at their promoters. These data suggest that, in addition to their importance in T2DM genetic risk, enhancers also govern important transcriptional regulatory changes accompanying or mediated by aging.

Circadian rhythm links behavior and metabolism to day–night cycles. Notably, insulin secretion oscillates with a circadian periodicity. Analysis of mouse islet transcriptomes revealed that approximately 27% of the beta cell transcriptome (*N*=3905 genes) demonstrated circadian oscillation, including genes responsible for insulin synthesis, transport, and stimulated exocytosis [50]. The human orthologs of 481 of these genes exhibited circadian oscillations in human islets. ChIP-seq identified 742 out of 3905 of these oscillatory genes as direct targets of the circadian clock TFs CLOCK and BMAL1. As with aging, most differential sites were at distal REs. Beta cell-specific deletion of *Bmal1* resulted in islet failure and diabetes in mice. This study demonstrates the importance of circadian rhythms in islet function and suggests that genetic or environmental perturbation of this program contribute to T2DM risk and pathophysiology. GWAS results suggest that this could be the case, because SNVs in the *CRY2* locus, a component of the circadian machinery, and *MTNR1B,* a gene encoding a melatonin receptor, are associated with altered islet function and T2DM [1,52]. It will be interesting to see whether genetic perturbations in circadian clock TFs or their binding sites emerge as one of the molecular mechanisms underlying T2DM GWAS.

Maternal nutrition and *in utero* stresses have been linked to T2DM risk for offspring in humans and rodents [23,53–55]. Although changes in fetal nutrition are suggested to influence offspring metabolism via epigenetic modifications [23,56], the genome-wide effects on the islet (epi) genome have not been determined. Similarly, stress responses to elevated oxidative and/or ER stress lead to islet failure, impaired insulin secretion, and T2DM susceptibility [57–59]. Ultimately, these responses converge on the nucleus and involve the redistribution or covalent modifications of master TFs (MAFB, NKX6-1, and PDX1) or stress response factors (FOXO1, ATF4, and HIF1alpha) [20,22,53,57,58]. (Epi)genomic analyses of these stress responses are warranted and may reveal important connections between T2DM SNVs and altered islet stress

responses. Moving forward, it will be crucial to understand the extent to which genetic and epigenetic changes interact in T2DM pathogenesis (see Outstanding Questions). Response QTL (reQTL) and epigenome-wide association studies (EWAS) [56] should provide these important links (see Outstanding Questions). Indeed, studies of SNV effects on immune cell responses identified 121 reQTLs, 38 of which overlapped autoimmune disease-associated SNVs [60]. Specific factor(s) and pathway(s) activated by insulin resistance appear to differ between mouse and human islets [57,58] (Figure 2); thus, we emphasize that caution must be taken in study design and interpretation to interrogate this and possibly other islet responses.

Deconstructing Pancreatic Islet Cellular and/or Functional Heterogeneity

Islets comprise 1–5% of the pancreas and consist of at least five endocrine cell types performing coordinated but distinct functions and each producing a unique hormone in the islet: beta (insulin), alpha (glucagon), delta (somatostatin), gamma (pancreatic polypeptide), and epsilon (ghrelin) cells [61–64]. Precise understanding of islet molecular changes during T2DM development is likely complicated by variability in islet cell type composition. On average, islets comprise 55% beta cells, 35% alpha, 10% delta, and less than 5% and 1% gamma/PP and epsilon cells, respectively [61–63]. However, this can vary considerably between donors, with ranges of 28.4–76.2%, 23.8–71.6%, and 2.4–12% for beta, alpha, and delta cell compositions, respectively [61] (Figure 2). This cellular heterogeneity, combined with donor-to-donor variability, masks the molecular repertoire of each cell type and impedes clear understanding of the molecular programs perturbed in each cell type by T2DM pathogenesis.

Until recently, most studies had focused on epigenetic and transcriptional analyses of whole islets or, at the expense of other cell types, beta cells. However, recent studies demonstrating roles for alpha [65–67] and delta cells [68–71] in modulating beta cell function and/or resilience and in T2DM pathogenesis are fueling renewed interest in these cell types. First attempts to overcome these obstacles and understand the molecular repertoire of each islet cell type focused on transcriptomic analyses of sorted and enriched cell type populations [61,72–74]. However, such methods were unable to effectively isolate and enrich the less abundant nonbeta cells [75], leaving much of the functional genomic landscape of islets imprecisely assigned and/or classified or, in the case of rarer islet cell types, undefined.

Within the past year, multiple groups have applied single cell transcriptome profiling to islets to begin to address questions about islet physiology [75–83] (see Outstanding Questions) with single cell resolution, such as: (i) what is the gene repertoire of each islet cell type? (ii) Does the gene repertoire reveal any new and/or unexpected roles for each cell type in islet (patho) physiology? (iii) Are there novel cell types or unappreciated **subpopulations** in islets? These studies are providing new appreciation of the repertoire of both islet beta and nonbeta cells. Given that much of the beta cell transcriptional repertoire has been extensively studied [61,72–74], several features have been validated, including genes involved in cell survival and/or maturation (*PDX1*), regulation of insulin secretion (*RGS16, SYT13,* and *ENTPD3*), and diabetes-associated genes (*DLK1, MEG3,* and *SLC2A2*) [75,76,78–81,83]. Unique expression of genes encoding TFs (*IRX2*), membrane glycoproteins (*DPP4*), and hormone transporters (*TTR*) were also validated in alpha cells. Analysis of single alpha cell transcriptomes uncovered signatures involved in wound healing (*FAP*), blood clotting (*F10*), and tissue biogenesis (*LOXL4*) [75,76,78–81,83], suggesting that they share functions akin to pancreatic fibroblast and/or mesenchymal cells.

Single cell profiling has provided new views of the roles of delta and PP/gamma cells in islet physiology and the molecular genetics of islet failure and diabetes. For example, these studies revealed that delta cells uniquely express appetite-suppressing leptin (*LEPR*) and appetite-stimulating ghrelin (*GHSR*) hormone receptors [75,79,80], implicating them as the integrators

and regulators of these pathways in the islet. *GHSR* functionality has been demonstrated in both human and mouse delta cells [70]. *LEPR* expression is unique to human delta cells, suggesting that these cells uniquely mediate the leptin response in human islets [70,75,76,79,80] (Figure 2). Expression of genes associated with congenital hyperinsulinemia (CHI) (*UCP2* and *HADH*) in delta cells further implicates this cell type in the molecular genetics of CHI [75]. PP/gamma cell transcriptomes exhibited enrichment of genes involved in neuronal development (*MEIS2* and *FEV*) [75,78–80] and serotonin catalysis and reuptake (*TPH1* and *SLC6A4*) [75,79,80,83]. Together, these findings suggest that delta and PP/gamma cells act as the 'brains' of the pancreatic islets, capable of receiving and integrating various neuronal signals to coordinate islet function. Due to their scarcity in human pancreatic islets (<1% of islet volume), our knowledge of the epsilon cell repertoire and its putative function(s) remain speculative. Nonetheless, the insights gleaned from these initial studies undoubtedly motivate follow-up studies that continue transitioning from whole-islet to functional constituent cell studies. Identification of genes encoding cell type-specific surface markers (beta, *LRRTM3* and *CASR*; alpha, *DPP4* and *PLCE1*; delta, *LEPR*, *GHSR*, and *ERBB4*; PP/gamma, *SLC6A4* and *PTGFR*; and epsilon, *ANXA13*) [75,79] provide new targets that may be exploited for more accurate purification of each islet cell type and analysis of its specific responses to genetic and environmental stressors.

Islet Subpopulations and Cell Type Heterogeneity

Detection of heterogeneous beta cell subpopulations was reported for enriched cell and single cell studies. These include four subpopulations with differing expression of *ST8SIA1* and *CD9* [84], five subpopulations defined by *RBP4, FFAR4/GPR120, ID1, ID2*, and *ID3* expression [80], and subpopulations characterized by ER stress-associated [76] and oxidative stress-associated genes [79]. *Fltp/CFAP126* expression has been reported to distinguish proliferating and mature beta cell subpopulations in mice [85], but single cell transcriptome analyses failed to detect this distinction in human beta cells [75,83]. However, proliferative and mature human beta cells were identified by single cell mass cytometry analysis [86], suggesting that mice and humans make use of distinct cell growth pathways. Given that each study detected distinct beta cell subpopulations with different gene signatures, it remains difficult to distinguish whether these subpopulations are functionally distinct cells or the result of technical confounders, such as the time to sort and enrich in a harsh cell sorting environment, time for cell capture, or cell and transcript capture efficiency [87].

Single Cell Dissection of Islet Dysfunction and T2DM

Single cell transcriptome analyses provide a fresh and agnostic opportunity to investigate the putative mechanisms underlying islet dysfunction in T2DM. To date, single cell transcriptome profiling has been completed for a total of 1831 and 1970 islet cells from 26 ND and 15 T2DM donors, respectively [75,80,81,83]. Comparison of T2DM and ND single cell transcriptomes suggest that specific alterations in islet cell type transcriptomes underlie T2DM pathogenesis (Figure 3A). However, changes in cell proportions (Figure 3B), identity, and plasticity (Figure 3C, D) have also been regarded as potential contributors to T2DM [72,88–92]. Specifically, decreases in diabetic beta cell mass were suggested to be caused by reversion to endocrine progenitor (hormone-negative) cells (Figure 3C) or different islet cell types (Figure 3D) rather than to apoptosis. The model of transformed beta cell identity remains controversial. A recent study concluded that the observed magnitude of decline in beta cell numbers in T2DM islets is not accompanied by proportionate increases in cells exhibiting ***trans*-differentiation** markers or increases in other islet cell types [93]. Rather, the presence of endocrine progenitor-like cells in T2DM islets may represent newly forming endocrine cells [93]. Single cell profiling also did not identify transcriptomic evidence of **dedifferentiated** or *trans*-differentiated cells in T2DM islets (Figure 3C,D) [75,80,83]. Similar trends were observed in whole-islet RNA-seq data upon **deconvolution**, where cell type proportions did not significantly vary between hypoglycemic

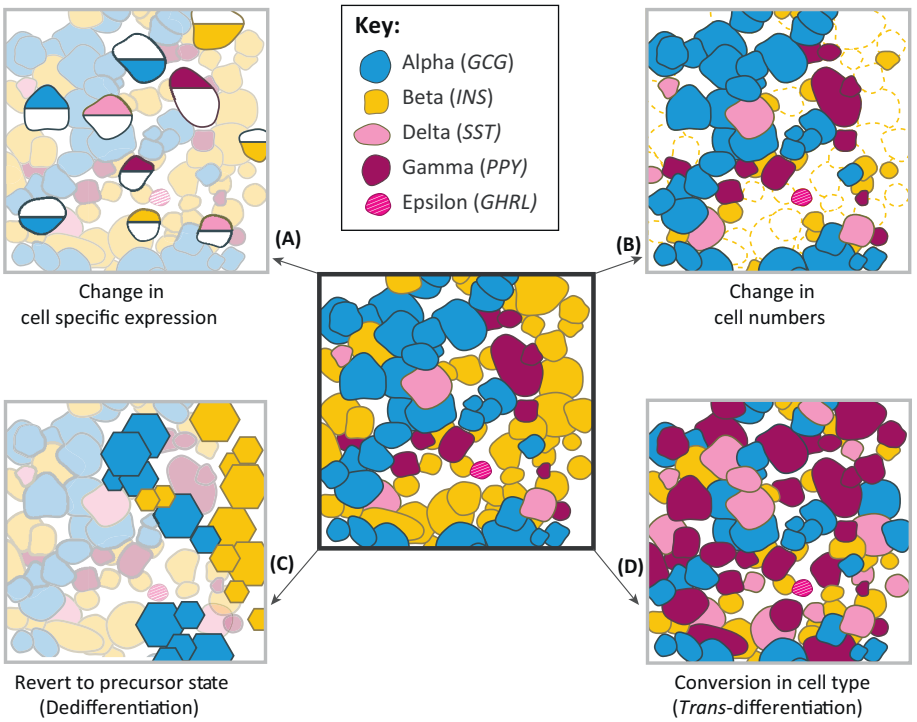


Figure 3. Proposed Cellular Mechanisms Contributing to Type 2 Diabetes Mellitus (T2DM) Development. (Center) Cartoon representation of human islet cellular composition. Studies have described the following phenomena: (A) Islet single cell transcriptomic studies [75,80,83] suggest that cell type-specific changes in gene expression (depicted as half-shaded cells) contribute to T2DM pathogenesis. These studies suggest that potential pathogenic expression changes occur in each islet cell type, not just beta cells. (B) Decreases in beta cell (in yellow) numbers [25,92,100,101], thought to precede islet dysfunction and development of insulin resistance. (C) Alterations in islet cellular identity may also account for islet failure. Dedifferentiation of islet cell types to precursor cell types and/or states (hexagons) has been proposed to underlie the loss of beta cell mass and function in T2DM [88,90–92]. (D) Similarly, *trans*-differentiation of islet cell types may lead to imbalances in islet cell proportions and improper function [72,88,89]

and hyperglycemic islets [76]. Thus, the transcriptome data to date do not provide supporting evidence of dedifferentiation in T2DM islets.

Transcriptomes of each cell type from ND and T2DM donors exhibited remarkable correlation overall. However, specific changes in gene expression were reported in T2DM beta cells, including reduced expression of *INS* [75,80], genes important for insulin secretion (*STX1A*) [75] and beta cell proliferation (*FXYD2*) [80,83], as well as elevated expression of genes implicated in T2DM GWAS (*DLK* and , *DGKB*) [75]. Transcriptional differences were also identified in T2DM alpha cells, including expression of *CD36* [75,80], a crucial activator of the NLRP3 inflammasome [94], and *RGS4,* a negative regulator of GSIS [80]. Several genes were dysregulated in T2DM delta cell transcriptomes [75,83]. However, the underlying biology of these candidates remains undefined, with no association with islet growth or function [83]. Aside from these encouraging examples, these single cell studies have not reached consensus regarding differentially expressed genes between T2DM and ND cell types. Differences in islet donor variability, islet isolation and/or transport, and single cell dissociation and/or sequencing protocols may explain these inconsistencies across studies. We expect that sampling thousands of single cells each from hundreds of individuals for large-scale meta-analyses will provide a more convergent list of cell type-specific genes and pathways disrupted in T2DM islets. It will also be important for future studies to profile cells from individuals at different points

along the T2DM pathogenesis spectrum, such as prediabetic individuals ($5.5<HbA1_c<6.0$) to identify and distinguish primary from secondary genomic changes that may be the cause or consequence of progression to T2DM.

Concluding Remarks and Future Directions

The past few years have marked exciting developments in our understanding of the underlying genomic, environmental, and cellular components driving T2DM pathogenesis. Numerous common (and only few rare) genetic SNVs have been implicated in T2DM progression [13,14]. It is unclear whether the 'missing T2DM heritability' is explained by a larger distribution of common SNVs with minimal effect sizes, whether current methods have missed critical rare SNVs, or whether it will be captured by gene–gene and gene–environment interactions (such as detected by reQTL). Thus far, most catalogued T2DM-SNVs occur in, and disrupt, islet RE function; however, the causal connections between the two remain challenging to decipher. eQTL and chromatin accessibility QTL (caQTL) [95,96] studies have been, and will continue to be, essential for linking genetic variants to molecular phenotypes. A subsequent challenge will be to link these molecular effects to pathways [39] and (patho)physiological phenotypes [97].

Functional genomic studies have identified minimal overlap between islet eQTLs and T2DM-SNVs [11,31], suggesting that responses to environmental stress factors are key mediators of T2DM pathogenesis. Mouse models have been instrumental in elucidating the genetic and molecular regulation of these responses and how environmental stressors influence islet (dys) function. However, observed differences between mice and humans in islet morphology, composition, expression, and function remind us to exercise caution when extrapolating findings in mice to human T2DM. Studies comparing the genomic features of human islets and models are essential to define conserved features and those that require modification to determine what aspects of islet dysfunction and T2DM we can model effectively and to decide how and/or where we should manipulate or humanize the mouse (epi)genome to better model human T2DM. (Epi)genome editing technologies, such as CRISPR/Cas9, can then be applied to develop new cellular and animal models to more effectively study islet phenotypic changes resulting from genetic and environmental variation. We anticipate that these integrative genomic studies and techniques will also serve as valuable resources to determine the underlying genetic changes and mechanisms of beta cell dysfunction that lead to T1DM [98].

Rapid developments in single cell NGS technologies have renewed interest in the less-studied islet cell types. Deconstructing the major molecular changes that occur in each cell type during T2DM progression has proven challenging, yielding inconsistent results between studies due to patient donor variability and technical sequencing artifacts. This is also likely the result of limited statistical power. In the future, it will be interesting to perform meta-analyses of available transcriptomic data to maximize our confidence of changes in cell specific expression programs. Together, the innovative new genomic technologies of the past few years will allow us to more precisely define, model, and manipulate the genes and pathways that have gone awry in T2DM, with the ultimate goal of designing novel therapeutic approaches.

Outstanding Questions

T2DM-associated GWAS variants explain only a small portion of T2DM heritability, with rare variants showing minimal contribution. Does a long tail of common variants with small effect sizes explain this missing heritability? Or are we simply 'underpowered' to detect rare variants and their contribution to T2DM heritability?

What are the genes targeted by T2DM GWAS sequence variant (SV)-containing regulatory elements? Are these links context specific? Does the risk allele enhance (gain-of-function) or repress (loss-of-function) RE function?

How do the transcriptomes and/or epigenomes of islets and islet cell types change when subjected to variable environmental stressors (oxidative stress, inflammation, diet, etc.)? How are they changed by intrinsic (aging, circadian rhythms, etc.) environmental factors? Which SVs regulate and alter these islet responses?

What are the precise cellular and molecular pathophysiological changes in each cell type that lead to T2DM? Are the major pathological changes beta cell specific or do they involve other islet cell types and/or non-islet cell types?

How many islet and single cell samples must be obtained to effectively capture combined cell type heterogeneity while controlling for technical and experimental confounders? How many samples are needed to observe genetic and/or epigenetic differences between T2DM and ND states? Would stratification of islets by T2DM risk genotype improve cell type-specific T2DM signatures?

Acknowledgments

Work in the Stitzel Lab is supported by the Assistant Secretary of Defense for Health Affairs, through the Peer Reviewed Medical Research Program under Award No. W81XWH-16-1-0130 and by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) under award number R00DK092251. Opinions, interpretations, conclusions, and recommendations are solely the responsibility of the authors, are not necessarily endorsed by the Department of Defense, and do not necessarily represent the official views of the National Institutes of Health. We thank Jane Cha for her expertise and help in figure design and artwork. We gratefully acknowledge members of the Stitzel and Ucar labs for helpful discussion and feedback on this work, and we thank the anonymous reviewers for their helpful suggestions to improve the perspective and content of this manuscript.

Supplemental Information

Supplemental Information associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.tig.2017.01.010

References

1. Mohlke, K.L. and Boehnke, M. (2015) Recent advances in understanding the genetic architecture of type 2 diabetes. *Hum. Mol. Genet.* 24, R85–R92
2. Parker, S.C.J. *et al.* (2013) Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. U. S. A.* 110, 17921–17926
3. Pasquali, L. *et al.* (2014) Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* 46, 136–143
4. Roadmap Epigenomics Consortium *et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330
5. Quang, D.X. *et al.* (2015) Motif signatures in stretch enhancers are enriched for disease-associated genetic variants. *Epigenetics Chromatin* 8, 23
6. Trynka, G. *et al.* (2013) Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.* 45, 124–130
7. Finucane, H.K. *et al.* (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* 47, 1228–1235
8. Stitzel, M.L. *et al.* (2015) Transcriptional regulation of the pancreatic islet: implications for islet function. *Curr. Diab. Rep.* 15, 66
9. Kulzer, J.R. *et al.* (2014) A common functional regulatory variant at a type 2 diabetes locus upregulates ARAP1 expression in the pancreatic beta cell. *Am. J. Hum. Genet.* 94, 186–197
10. Fogarty, M.P. *et al.* (2014) Identification of a regulatory variant that binds FOXA1 and FOXA2 at the CDC123/CAMK1D type 2 diabetes GWAS locus. *PLOS Genet.* 10, e1004633
11. van de Bunt, M. *et al.* (2015) Transcript expression data from human islets links regulatory signals from genome-wide association studies for type 2 diabetes and glycemic traits to their downstream effectors. *PLoS Genet.* 11, e1005694
12. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium *et al.* (2014) Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* 46, 234–244
13. Morris, A.P. *et al.* (2012) Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* 44, 981–990
14. Fuchsberger, C. *et al.* (2016) The genetic architecture of type 2 diabetes. *Nature* 536, 41–47
15. Gaulton, K.J. *et al.* (2015) Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat. Genet.* 47, 1415–1425
16. Agarwala, V. *et al.* (2013) Evaluating empirical bounds on complex disease genetic architecture. *Nat. Genet.* 45, 1418–1427
17. Strawbridge, R.J. (2011) Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* 60, 2624–2634
18. Steinthorsdottir, V. (2014) Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat. Genet.* 46, 294–298
19. Hara, K. *et al.* (2014) Genome-wide association study identifies three novel loci for type 2 diabetes. *Hum. Mol. Genet.* 23, 239–246
20. Aouacheri, O. *et al.* (2015) The investigation of the oxidative stress-related parameters in type 2 diabetes mellitus. *Can. J. Diabetes* 39, 44–49
21. Chaudhari, N. *et al.* (2014) A molecular web: endoplasmic reticulum stress, inflammation, and oxidative stress. *Front. Cell. Neurosci.* 8, 213
22. Gorasia, D.G. *et al.* (2015) Pancreatic beta cells are highly susceptible to oxidative and ER stresses during the development of diabetes. *J. Proteome Res.* 14, 688–699
23. Rando, O.J. and Simmons, R.A. (2015) I'm eating for two: parental dietary effects on offspring metabolism. *Cell* 161, 93–105
24. Zephy, D. and Ahmad, J. (2015) Type 2 diabetes mellitus: role of melatonin and oxidative stress. *Diabetes Metab. Syndr. Clin. Res. Rev.* 9, 127–131
25. Laybutt, D.R. *et al.* (2007) Endoplasmic reticulum stress contributes to beta cell apoptosis in type 2 diabetes. *Diabetologia* 50, 752–763
26. Scott, L.J. *et al.* (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316, 1341–1345
27. Sladek, R. *et al.* (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445, 881–885
28. Burton, P.R. *et al.* (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661–678
29. Diabetes Genetics Initiative of the Broad Institute of Harvard *et al.* (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316, 1331–1336
30. Liu, C.-T. *et al.* (2016) Trans-ethnic meta-analysis and functional annotation illuminates the genetic architecture of fasting glucose and insulin. *Am. J. Hum. Genet.* 99, 56–75
31. Fadista, J. *et al.* (2014) Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. *Proc. Natl. Acad. Sci. U. S. A.* 111, 13924–13929
32. Stitzel, M.L. *et al.* (2010) Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell Metab.* 12, 443–455
33. Ackermann, A.M. *et al.* (2016) Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol. Metab.* 5, 233–244
34. Hnisz, D. *et al.* (2013) Super-enhancers in the control of cell identity and disease. *Cell* 155, 934–947
35. Dekker, J. and Mirny, L. (2016) The 3D genome as moderator of chromosomal communication. *Cell* 164, 1110–1121
36. Lieberman-Aiden, E. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293
37. Li, G. *et al.* (2010) ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol.* 11, R22
38. Mumbach, M.R. *et al.* (2016) HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* 13, 919–922
39. Keller, M.P. *et al.* (2016) The transcription factor Nfatc2 regulates β-cell proliferation and genes associated with type 2 diabetes in mouse and human islets. *PLOS Genet.* 12, e1006466
40. Scott, L.J. *et al.* (2016) The genetic regulatory signature of type 2 diabetes in human skeletal muscle. *Nat. Commun.* 7, 11764
41. Small, K.S. *et al.* (2011) Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat. Genet.* 43, 561–564
42. Morán, I. *et al.* (2012) Human β cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. *Cell Metab.* 16, 435–448
43. Arnes, L. and Sussel, L. (2015) Epigenetic modifications and long noncoding RNAs influence pancreas development and function. *Trends Genet.* 31, 290–299
44. Voight, B.F. *et al.* (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* 42, 579–589

45. Wallace, C. *et al.* (2010) The imprinted DLK1-MEG3 gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes. *Nat. Genet.* 42, 68–71
46. Dayeh, T. *et al.* (2014) Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genet.* 10, e1004160
47. Volkmar, M. *et al.* (2012) DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients. *EMBO J.* 31, 1405–1426
48. Arda, H.E. *et al.* (2016) Age-dependent pancreatic gene regulation reveals mechanisms governing human β cell function. *Cell Metab.* 23, 909–920
49. Avrahami, D. *et al.* (2015) Aging-dependent demethylation of regulatory elements correlates with chromatin state and improved β cell function. *Cell Metab.* 22, 619–632
50. Perelis, M. *et al.* (2015) Pancreatic β cell enhancers regulate rhythmic transcription of genes controlling insulin secretion. *Science* 350, aac4250
51. Kluth, O. *et al.* (2014) Differential transcriptome analysis of diabetes-resistant and -sensitive mouse islets reveals significant overlap with human diabetes susceptibility genes. *Diabetes* 63, 4230–4238
52. Persaud, S.J. and Jones, P.M. (2016) A wake-up call for type 2 diabetes? *N. Engl. J. Med.* 375, 1090–1092
53. Halban, P.A. *et al.* (2014) β-cell failure in type 2 diabetes: postulated mechanisms and prospects for prevention and treatment. *Diabetes Care* 37, 1751–1758
54. Schulz, L.C. (2010) The Dutch Hunger Winter and the developmental origins of health and disease. *Proc. Natl. Acad. Sci.* 107, 16757–16758
55. Li, J. *et al.* (2017) Prenatal exposure to famine and the development of hyperglycemia and type 2 diabetes in adulthood across consecutive generations: a population-based cohort study of families in Suihua, China. *Am. J. Clin. Nutr.* 105, 221–227
56. Rakyan, V.K. *et al.* (2011) Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.* 12, 529–541
57. Dai, C. *et al.* (2016) Stress-impaired transcription factor expression and insulin secretion in transplanted human islets. *J. Clin. Invest.* 126, 1857–1870
58. Dai, C. *et al.* (2012) Islet-enriched gene expression and glucose-induced insulin secretion in human and mouse islets. *Diabetologia* 55, 707–718
59. Guo, S. *et al.* (2013) Inactivation of specific β cell transcription factors in type 2 diabetes. *J. Clin. Invest.* 123, 3305–3316
60. Lee, M.N. *et al.* (2014) Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science* 343, 1246980
61. Blodgett, D.M. *et al.* (2015) Novel observations from next-generation RNA sequencing of highly purified human adult and fetal islet cell subsets. *Diabetes* 64, 3172–3181
62. Brissova, M. *et al.* (2005) Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy. *J. Histochem. Cytochem.* 53, 1087–1097
63. Cabrera, O. *et al.* (2006) The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. *Proc. Natl. Acad. Sci. U. S. A.* 103, 2334–2339
64. Ionescu-Tirgoviste, C. (2015) A 3D map of the islet routes throughout the healthy human pancreas. *Sci. Rep.* 5, 14634
65. Stanojevic, V. and Habener, J.F. (2015) Evolving function and potential of pancreatic alpha cells. *Best Pract. Res. Clin. Endocrinol. Metab.* 29, 859–871
66. Jamison, R.A. *et al.* (2011) Hyperglucagonemia precedes a decline in insulin secretion and causes hyperglycemia in chronically glucose-infused rats. *Am. J. Physiol. Endocrinol. Metab.* 301, E1174–E1183
67. Rodriguez-Diaz, R. (2011) Alpha cells secrete acetylcholine as a non-neuronal paracrine signal priming beta cell function in humans. *Nat. Med.* 17, 888–892
68. Hauge-Evans, A.C. (2009) Somatostatin secreted by islet delta-cells fulfills multiple roles as a paracrine regulator of islet function. *Diabetes* 58, 403–411
69. van der Meulen, T. (2015) Urocortin3 mediates somatostatin-dependent negative feedback control of insulin secretion. *Nat. Med.* 21, 769–776
70. DiGruccio, M.R. *et al.* (2016) Comprehensive alpha, beta and delta cell transcriptomes reveal that ghrelin selectively activates delta cells and promotes somatostatin release from pancreatic islets. *Mol. Metab.* 5, 449–458
71. Molina, J. *et al.* (2014) Control of insulin secretion by cholinergic signaling in the human pancreatic islet. *Diabetes* 63, 2714–2726
72. Bramswig, N.C. *et al.* (2013) Epigenomic plasticity enables human pancreatic α to β cell reprogramming. *J. Clin. Invest.* 123, 1275–1284
73. Dorrell, C. *et al.* (2011) Transcriptomes of the major human pancreatic cell types. *Diabetologia* 54, 2832–2844
74. Nica, A.C. *et al.* (2013) Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. *Genome Res.* 23, 1554–1562
75. Lawlor, N. *et al.* (2017) Single cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. *Genome Res.* 27, 208–222
76. Baron, M. *et al.* (2016) A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Syst.* 3, 346–360
77. Grün, D. *et al.* (2016) De novo prediction of stem cell identity using single-cell transcriptome data. *Cell Stem Cell* 19, 266–277
78. Li, J. *et al.* (2016) Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types. *EMBO Rep.* 17, 178–187
79. Muraro, M.J. *et al.* (2016) A single-cell transcriptome atlas of the human pancreas. *Cell Syst.* 3, 385–394
80. Segerstolpe, Å. *et al.* (2016) Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell Metab.* 24, 593–607
81. Wang, Y.J. *et al.* (2016) Single cell transcriptomics of the human endocrine pancreas. *Diabetes* 65, 3028–3038
82. Xin, Y. *et al.* (2016) Use of the Fluidigm C1 platform for RNA sequencing of single mouse pancreatic islet cells. *Proc. Natl. Acad. Sci. U. S. A.* 113, 3293–3298
83. Xin, Y. *et al.* (2016) RNA sequencing of single human islet cells reveals type 2 diabetes genes. *Cell Metab.* 24, 608–615
84. Dorrell, C. *et al.* (2016) Human islets contain four distinct subtypes of β cells. *Nat. Commun.* 7, 11756
85. Bader, E. *et al.* (2016) Identification of proliferative and mature β-cells in the islets of Langerhans. *Nature* 535, 430–434
86. Wang, Y.J. *et al.* (2016) Single-cell mass cytometry analysis of the human endocrine pancreas. *Cell Metab.* 24, 616–626
87. Liu, S. and Trapnell, C. (2016) Single-cell transcriptome sequencing: recent advances and remaining challenges. *F1000Research* 5, 182
88. Cinti, F. *et al.* (2016) Evidence of β-cell dedifferentiation in human type 2 diabetes. *J. Clin. Endocrinol. Metab.* 101, 1044–1054
89. Lu, J. *et al.* (2014) Transdifferentiation of pancreatic α-cells into insulin-secreting cells: from experimental models to underlying mechanisms. *World J. Diabetes* 5, 847–853
90. Talchai, C. *et al.* (2012) Pancreatic β-cell dedifferentiation as mechanism of diabetic β-cell failure. *Cell* 150, 1223–1234
91. Wang, Z. *et al.* (2014) Pancreatic β-cell dedifferentiation in diabetes and re-differentiation following insulin therapy. *Cell Metab.* 19, 872–882
92. Butler, A.E. *et al.* (2003) Beta-cell deficit and increased beta-cell apoptosis in humans with type 2 diabetes. *Diabetes* 52, 102–110
93. Butler, A.E. *et al.* (2016) β-cell deficit in obese type 2 diabetes, a minor role of β-cell dedifferentiation and degranulation. *J. Clin. Endocrinol. Metab.* 101, 523–532
94. Sheedy, F.J. *et al.* (2013) CD36 coordinates NLRP3 inflammasome activation by facilitating the intracellular nucleation

from soluble to particulate ligands in sterile inflammation. *Nat. Immunol.* 14, 812–820

95. Kumasaka, N. *et al.* (2016) Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nat. Genet.* 48, 206–213
96. Degner, J.F. *et al.* (2012) DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* 482, 390–394
97. Thomsen, S.K. *et al.* (2016) Systematic functional characterization of candidate causal genes for type 2 diabetes risk variants. *Diabetes* 65, 3805–3811
98. Soleimanpour, S.A. and Stoffers, D.A. (2013) The pancreatic β cell and type 1 diabetes: innocent bystander or active participant? *Trends Endocrinol. Metab.* 24, 324–331
99. Maller, J.B. *et al.* (2012) Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* 44, 1294–1301
100. Cerf, M.E. (2013) Beta cell dysfunction and insulin resistance. *Front. Endocrinol.* 4, 37
101. Halban, P.A. *et al.* (2014) β-cell failure in type 2 diabetes: postulated mechanisms and prospects for prevention and treatment. *Diabetes Care* 37, 1751–1758

Single-cell transcriptomes identify human islet cell signatures and reveal cell-type–specific expression changes in type 2 diabetes

Nathan Lawlor,[1,4] Joshy George,[1,4] Mohan Bolisetty,[1] Romy Kursawe,[1] Lili Sun,[1] V. Sivakamasundari,[1] Ina Kycia,[1] Paul Robson,[1,2,3] and Michael L. Stitzel[1,2,3]

[1] *The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut 06032, USA;* [2] *Institute for Systems Genomics, University of Connecticut, Farmington, Connecticut 06032, USA;* [3] *Department of Genetics & Genome Sciences, University of Connecticut, Farmington, Connecticut 06032, USA*

Blood glucose levels are tightly controlled by the coordinated action of at least four cell types constituting pancreatic islets. Changes in the proportion and/or function of these cells are associated with genetic and molecular pathophysiology of monogenic, type I, and type 2 (T2D) diabetes. Cellular heterogeneity impedes precise understanding of the molecular components of each islet cell type that govern islet (dys)function, particularly the less abundant delta and gamma/pancreatic polypeptide (PP) cells. Here, we report single-cell transcriptomes for 638 cells from nondiabetic (ND) and T2D human islet samples. Analyses of ND single-cell transcriptomes identified distinct alpha, beta, delta, and PP/gamma cell-type signatures. Genes linked to rare and common forms of islet dysfunction and diabetes were expressed in the delta and PP/gamma cell types. Moreover, this study revealed that delta cells specifically express receptors that receive and coordinate systemic cues from the leptin, ghrelin, and dopamine signaling pathways implicating them as integrators of central and peripheral metabolic signals into the pancreatic islet. Finally, single-cell transcriptome profiling revealed genes differentially regulated between T2D and ND alpha, beta, and delta cells that were undetectable in paired whole islet analyses. This study thus identifies fundamental cell-type–specific features of pancreatic islet (dys)function and provides a critical resource for comprehensive understanding of islet biology and diabetes pathogenesis.

[Supplemental material is available for this article.]

Pancreatic islets of Langerhans are clusters of at least four different hormone-secreting endocrine cell types that elicit coordinated—but distinct—responses to maintain glucose homeostasis. As such, they are central to diabetes pathophysiology. On average, human islets consist mostly of beta (54%), alpha (35%), and delta (11%) cells; up to a few percent gamma/pancreatic polypeptide (PP) cells; and very few epsilon cells (Brissova et al. 2005; Cabrera et al. 2006; Blodgett et al. 2015). Human islet composition is neither uniform nor static but varies between individuals and across regions of the pancreas (Brissova et al. 2005; Cabrera et al. 2006; Blodgett et al. 2015). Cellular heterogeneity complicates molecular studies of whole human islets and may mask important role(s) for less common cells in the population (Dorrell et al. 2011b; Bramswig et al. 2013; Nica et al. 2013; Blodgett et al. 2015; Liu and Trapnell 2016). Moreover, it complicates attempts to identify epigenetic and transcriptional signatures distinguishing diabetic from nondiabetic (ND) islets, leading to inconsistent reports of genes and pathways affected (Gunton et al. 2005; Marselli et al. 2010; Taneera et al. 2012; Dayeh et al. 2014). Conventional sorting and enrichment techniques are unable to specifically purify each human islet cell type (Dorrell et al. 2008; Nica et al. 2013; Bramswig et al. 2013; Hrvatin et al. 2014; Blodgett et al. 2015), thus a precise understanding of the transcriptional repertoire governing each cell type's identity and function is lacking. Identifying the cell-type–specific expression programs that contribute to islet dysfunction and type 2 diabetes (T2D) should reveal novel targets and approaches to prevent, monitor, and treat T2D.

In this study, we sought to decipher the transcriptional repertoire of each islet cell type in an agnostic and precise manner by capturing and profiling pancreatic single cells from ND and T2D individuals. From these profiles, we identified transcripts uniquely important for each islet cell type's identity and function. Finally, we compared T2D and ND individuals to identify islet cell-type–specific expression changes that were otherwise masked by islet cellular heterogeneity. The insights and data from this study provide an important foundation to guide future genomics-based interrogation of islet dysfunction and diabetes.

Results

Islet single-cell transcriptomes accurately recapitulate those of intact islets

Pancreatic islets (>85% purity and >90% viability) were obtained from eight human cadaveric organ donors (five ND, three T2D) (Fig. 1A; Supplemental Table S1). Each islet sample was processed to generate single-cell RNA-seq libraries (Fig. 1A; single cell) and paired bulk RNA-seq libraries at three different stages of islet processing (Fig. 1A; baseline, intact, and dissociated). All RNA-seq

[4] **These authors contributed equally to this work.**
Corresponding author: michael.stitzel@jax.org
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.212720.116. Freely available online through the *Genome Research* Open Access option.

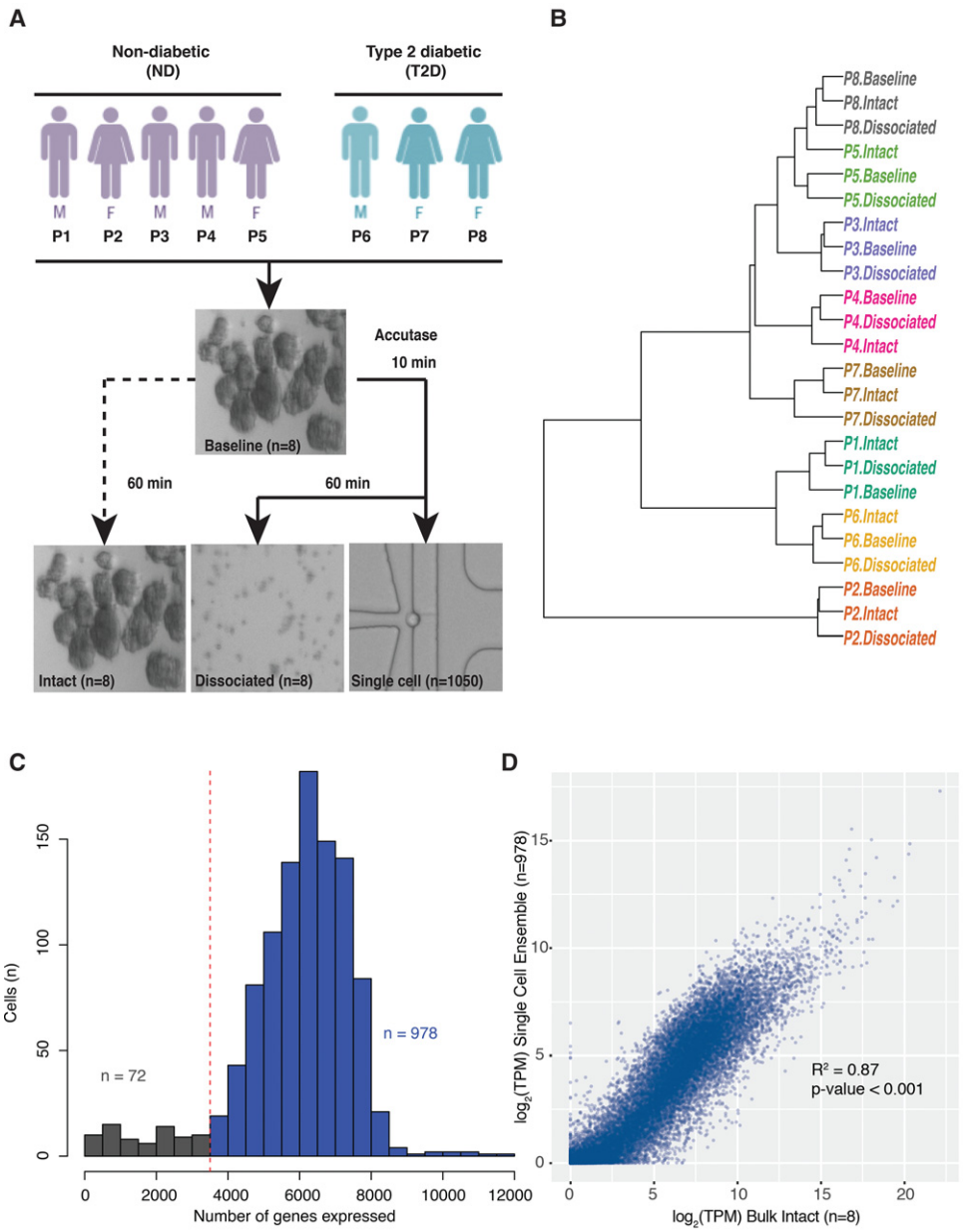


Figure 1. Single-cell transcriptomes reflect those of paired intact islets. (*A*) Schematic of experimental workflow. Islets from each donor sample ($n = 8$ individuals) were dissociated using Accutase, and single-cell transcriptomes were synthesized from 1050 cells captured using 11 Fluidigm C1 chips. In parallel, "bulk" RNA-seq libraries were prepared from remaining dissociated single cells (dissociated) and from intact islets either flash frozen (baseline) or incubated/processed (intact). (*B*) Unsupervised hierarchical clustering of baseline, intact, and dissociated islet transcriptomes demonstrates clustering by person and not by processing/experimental condition. (*C*) Histogram demonstrating the number of genes detected in each single cell. Cells expressing less than 3500 genes ($n = 72$) were removed from downstream analyses. (*D*) Scatter plot comparing intact islet bulk RNA-seq ($n = 8$) and ensemble single-cell RNA-seq ($n = 978$) data demonstrates high correlation. (R^2) Pearson's *R*-squared; (TPM) transcripts per million; (P) person.

methods employed SMARTer chemistry (Methods), and bulk islet cDNA libraries were sequenced to an average approximate depth of 34 million reads (Supplemental Table S2). Baseline, intact, and dissociated transcriptomes from each person were highly correlated (Supplemental Fig. S1). Transcriptomes clustered by donor and not by processing condition or incubation time (Fig. 1B), strongly suggesting that islet processing did not significantly alter islet transcriptomes.

A total of 1050 islet cells (622 ND and 428 T2D) were captured on 11 Fluidigm C1 chips. cDNA libraries were constructed from captured cells and barcoded, fragmented, pooled, and sequenced to an average depth of 3 million reads (Supplemental Table S2). Two separate library preparations from the same amplified cDNA for each of 83 single cells demonstrated remarkable correlation, suggesting minimal batch effects resulting from the cDNA processing and sequencing steps. Resequenced samples are highlighted in Supplemental Table S2 but were not included in subsequent analyses. Transcript coverage is indicated in Supplemental Figure S2. Approximately 81% (21,484/26,616) of protein-coding genes and long intergenic noncoding RNAs (lincRNAs) were detected

in at least one cell from the collection. On average, each single cell expressed 5944 genes (Fig. 1C). Cells expressing less than 3500 genes ($n = 72$) also exhibited high mitochondrial alignment rates and other reported transcriptional metrics of cell death and/or poor quality (Ilicic et al. 2016; Xin et al. 2016) and were removed from subsequent analyses (Fig. 1C).

We next assessed the extent to which the remaining 978 single-cell transcriptomes represent the expression patterns observed in intact islets. Single-cell transcriptome ensembles from each person were highly correlated (Pearson's R^2 ranged from 0.91–0.98) (Supplemental Fig. S3), regardless of disease state. Pearson's R^2 values between individuals' single-cell ensembles and corresponding "bulk" transcriptomes ranged from 0.75–0.86 (Supplemental Fig. S4) and did not differ substantially between ND ($R^2 = 0.87$) and T2D ($R^2 = 0.85$) samples (Supplemental Fig. S5). Overall, ensemble/aggregate single-cell transcriptome profiles correlated well with those of pooled bulk islet transcriptomes from all individuals (Fig. 1D, $R^2 = 0.87$). These results suggest that the islet single-cell transcriptomes are high quality and effectively reflect bulk islet transcriptomes.

Single-cell profiling captures transcriptomes of major and minor pancreatic endocrine and exocrine cell types

Five islet endocrine cell types have been assigned based on exclusive and robust expression of the peptide hormone genes *INS* (beta), *GCG* (alpha), *SST* (delta), *PPY* (PP/gamma), and *GHRL* (epsilon) (Baetens et al. 1979; Nussey and Whitehead 2001; Zhao et al. 2008; Li et al. 2016; Xin et al. 2016; Wang et al. 2016). The pancreas also contains three exocrine cell types—acinar, stellate, and ductal—that critically support digestion through synthesis and transport of digestive enzymes (Pandol 2011; Reichert and Rustgi 2011). Each also has been identified by specific marker gene expression, including the serine peptidase gene *PRSS1* (acinar) (Dabbs 2013), the extracellular matrix protein gene *COL1A1* (stellate) (Mathison et al. 2010), and the structural keratin gene *KRT19* (ductal) (Dorrell et al. 2008, 2011a,b; Reichert and Rustgi 2011). We used these marker genes to determine the representation of each islet cell type among our 978 profiled single cells.

Density plots (Fig. 2A) revealed bimodal expression of each marker gene across the population of single cells. Therefore, we employed Gaussian mixture modeling (GMM) to classify the cells unambiguously (Fig. 2B). Approximate $\log_2$ counts per million (CPM) thresholds for each marker gene used to classify cell types are provided in Supplemental Table S3. This approach identified 617 single cells (~63%) from T2D and ND islets expressing a single marker gene representative of each major endocrine and exocrine cell type, examples of which are shown in Figure 2C. This included 239 alpha, 264 beta, 25 delta, and 18 PP/gamma cells (Table 1); the proportions of each cell type are in the ranges previously reported (Brissova et al. 2005; Cabrera et al. 2006; Blodgett et al. 2015). Only one cell expressing high levels ($\log_2$CPM > 15) of *GHRL* was identified, which we presume to be an exceedingly rare epsilon cell. Additionally, we obtained 19 stellate, 24 acinar, and 27 ductal cells (Table 1), presumably from exocrine contamination of the islet cell preparations. Only 21 cells (~2%) expressed none of the specified marker genes (Table 1). Approximately one-third (340/978) of cells expressed more than one marker gene; these were removed from subsequent analysis due to concerns that these represent two vertically stacked cells in a given capture site (for details, see Methods). Similar ratios of potential stacked cells have been reported in other studies using the Fluidigm C1 platform to capture mouse (Xin et al. 2016) and human islet cells (Wang et al. 2016). Collectively, these single-cell data capture transcriptome profiles representing each of the major and minor pancreatic endocrine and exocrine cell types. Genome Browser tracks representing aggregate single-cell expression for each islet cell type have been generated using HOMER (Heinz et al. 2010) and are made available (see Data Access) to facilitate their use and investigation by members of the islet biology and diabetes research communities.

Unsupervised analyses of islet single-cell transcriptomes identify discrete clusters corresponding to cell type

To determine if and how islet cell transcriptomes cluster, we completed unsupervised dimensionality reduction via t-distributed stochastic neighbor embedding (t-SNE) on 380 ND single-cell samples (excluding "multiple" labeled samples). t-SNE assembled single-cell transcriptomes into discrete clusters based upon 1824 highly expressed genes (see Methods; Supplemental Table S4); GMM-based marker gene analysis revealed that each cluster corresponded to a distinct endocrine and exocrine cell type (Fig. 3A; Supplemental Fig. S6). Unsupervised hierarchical clustering also grouped single-cell transcriptomes into discrete cell types (Fig. 3B). Despite being obtained from different individuals, 161/168 beta, 128/138 alpha, 15/16 delta, and 12/12 PP/gamma cell transcriptomes clustered into the same dendrogram branches, strongly suggesting that cell type encodes the greatest variation in the data. Exocrine cells and those expressing none of the specified marker genes ("none") clustered separately from the endocrine cell types. Importantly, this clustering was driven by neither sequencing depth (Supplemental Fig. S7B) nor expression of classic marker genes (*INS*, *GCG*, *SST*, *PPY*, *GHRL*, *COL1A1*, *PRSS1*, and *KRT19*), as cells continued to cluster into discrete cell types even when all marker genes were removed from the expression data sets (Supplemental Figs. S7C, S8). Recent studies have reported heterogeneity among beta cells. Specifically, Dorrell et al. characterized four subpopulations of human beta cells based on differing *ST8SIA1* and *CD9* expression (Dorrell et al. 2016). Similarly, Bader et al. 2016 distinguished two populations of proliferating (*Fltp*$^+$) and mature (*Fltp*$^-$) mouse beta cells. We did not find evidence of beta cell subpopulations (Supplemental Fig. S9), nor did we identify numerous proliferating cells (Supplemental Table S5). T2D single-cell transcriptomes ($n = 258$) also demonstrated clear clustering by cell type in unsupervised analyses (Supplemental Figs. S10–S14) based on 1908 highly expressed genes (Supplemental Table S4). Thus, each endocrine and exocrine pancreatic cell type exhibits a complex characteristic expression signature that uniquely identifies it.

Differential expression analyses reveal islet cell-type–specific transcriptional signatures

To identify gene signatures distinguishing each islet cell type, we completed a series of pairwise differential expression analyses (Supplemental Table S6) between each cell type (see Methods). After intersecting the results from each pairwise comparison, we identified a conservative collection of 154 islet endocrine cell-type "signature" genes (61 beta, 51 alpha, 17 delta, 25 gamma), as well as 202 exocrine genes (109 stellate, 31 acinar, 62 ductal) at 5% false-discovery rate (FDR) (Fig. 3C; Supplemental Table S7). Two genes exhibited overlap between the endocrine and exocrine signature lists: *FAP* (alpha and stellate cell overlap) and *TNS1* (beta and stellate cell overlap). Gene set enrichment analysis (GSEA) identified enrichment (FDR-adjusted *P*-value <0.05) of insulin

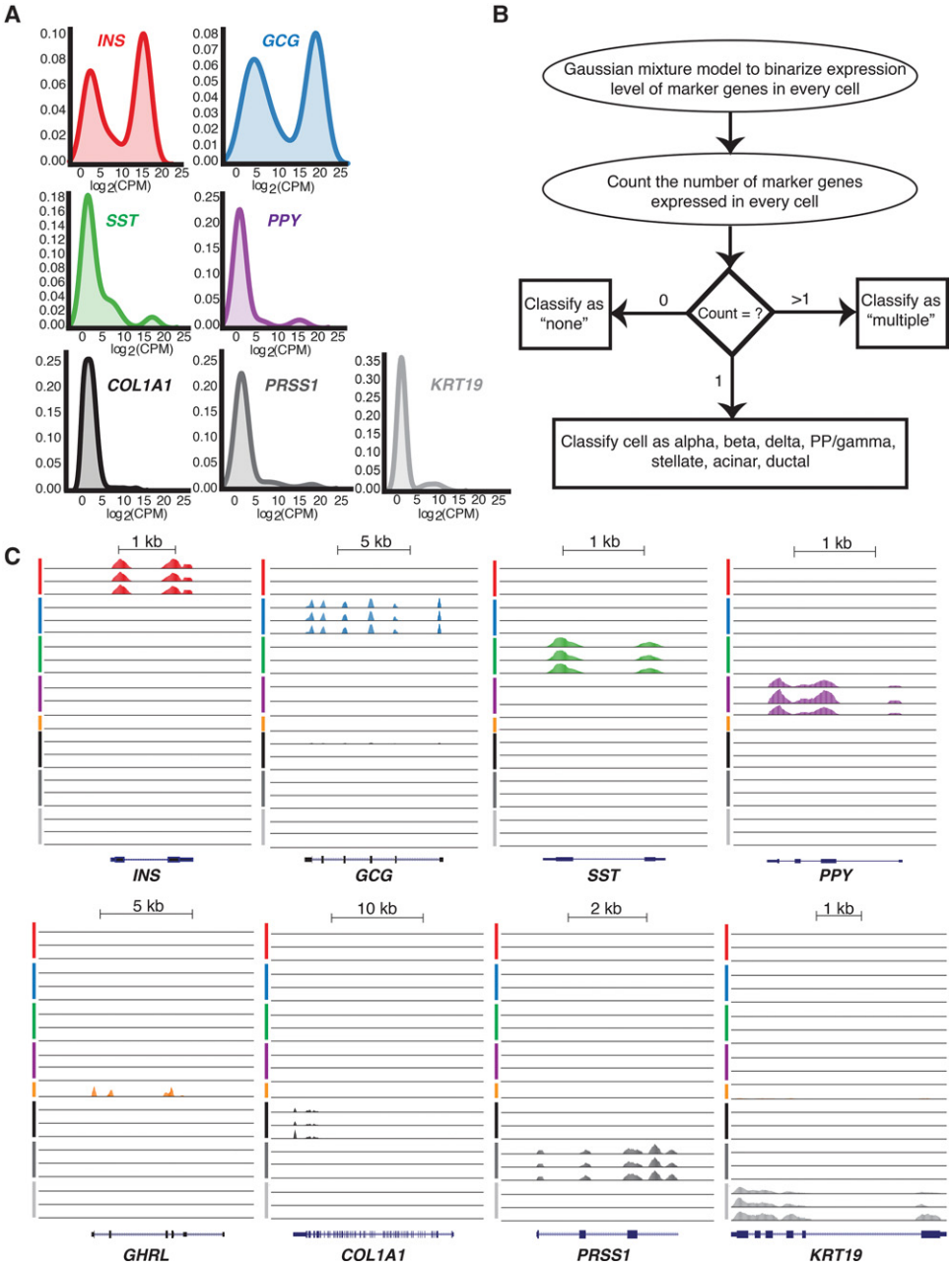


Figure 2. Cell-type classification based on marker gene expression. (*A*) Density plots demonstrating endocrine and exocrine marker gene expression across all single cells. (*B*) Schematic of the Gaussian mixture model method applied to assign cell-type identity based on marker gene expression. (*C*) UCSC Genome Browser views of representative single-cell expression profiles of *INS*, *GCG*, *SST*, *PPY*, and *GHRL* genes encoding beta, alpha, delta, PP/gamma, and epsilon cell hormones of the endocrine pancreas, respectively, and marker genes for stellate (*COL1A1*), acinar (*PRSS1*), and ductal (*KRT19*) cells of the exocrine pancreas. Line colors indicate putative beta (red), alpha (blue), delta (green), PP/gamma (purple), epsilon (orange), stellate (black), acinar (dark gray), and ductal cells (light gray). (PP) pancreatic polypeptide; (CPM) counts per million.

signaling, oxidative phosphorylation, maturity-onset diabetes of the young (MODY), and glycolysis/gluconeogenesis KEGG pathways in beta cells relative to the other endocrine cells (Supplemental Table S8).

Signature genes included previously reported beta-specific genes like *NKX6-1*, *DLK1*, and *ADCYAP1* (Fig. 3C, right) and alpha cell–specific genes like *IRX2*, *LOXL4*, and *DPP4*, a cell surface receptor and diabetes drug target (Dorrell et al. 2011a; Bramswig et al. 2013; Nica et al. 2013; Blodgett et al. 2015). Among delta cell signature genes, we detected exclusive expression of *HHEX*, a transcription factor reported to govern delta cell identity and function and linked to T2D GWAS (Zhang et al. 2014). Delta cells also specifically expressed *BCHE*, which encodes butyrylcholinesterase. BCHE catalyzes the breakdown of acetylcholine and ghrelin (Chen et al. 2015), thus providing a mechanism for delta cells to exert local inhibition of islet-influencing endocrine signals. PP/gamma cell–specific transcriptomes included *CTD-2008P7.8*, a lincRNA of unknown function; *CNTNAP5*, a member

Table 1. Number of profiled cells for each pancreatic cell type based on marker gene expression

Putative cell type (marker gene)	Cell ontology accession no.	Nondiabetic (ND)	Type 2 diabetic (T2D)
Alpha (*GCG*)	CL:0000171	138 (23.47%)	101 (25.9%)
Beta (*INS*)	CL:0000169	168 (28.57%)	96 (24.62%)
Delta (*SST*)	CL:0000173	16 (2.72%)	9 (2.31%)
PP/gamma (*PPY*)	CL:0002275	12 (2.04%)	6 (1.54%)
Epsilon (*GHRL*)	CL:0005019	1 (0.17%)	0
Stellate (*COL1A1*)	CL:0002410	9 (1.53%)	10 (2.56%)
Acinar (*PRSS1*)	CL:0002064	15 (2.55%)	9 (2.31%)
Ductal (*KRT19*)	CL:0002079	11 (1.87%)	16 (4.1%)
Multiple	—	208 (35.37%)	132 (33.85%)
None (other)	—	10 (1.7%)	11 (2.82%)
Total		588	390

of the neurexin family of cell adhesion molecules; and *ID4*, which encodes an inhibitor of DNA-binding protein. In addition to *DPP4*, we detected 30 islet signature genes whose proteins SWISSPROT predicts to localize to the cell surface (Supplemental Table S9). DPP4 antibodies have recently been used to isolate purer alpha cell populations from islets (Arda et al. 2016). Thus, antibodies against the other candidate cell-type–specific surface markers we have identified may be useful to purify other islet cell types.

Single-cell profiling identifies unexpected overlap in expression between minor and major islet cell types

Cell sorting and enrichment methods such as fluorescence-activated cell sorting (FACS) have been used to identify characteristic alpha and beta cell genes (Dorrell et al. 2011a,b; Bramswig et al. 2013; Nica et al. 2013; Blodgett et al. 2015). However, expression of *SST* or *PPY* in the reported alpha and beta cell gene sets suggests the presence of the less abundant delta and PP/gamma islet cell types in the enriched cell preparations. To distinguish genes exhibiting alpha- and beta-specific gene expression from those expressed also in delta and PP/gamma cells, we investigated the expression of previously reported alpha- and beta-specific genes (Supplemental Table S10; Supplemental Fig. S15) in our ND endocrine single-cell transcriptomes. Only 115/1683 previously reported beta-specific genes were expressed greater than fourfold higher in beta cells relative to the other endocrine cells (FDR < 0.05; one-way ANOVA followed by Tukey's honest significant difference [THSD]) (Fig. 3D). Similarly, 75/1853 reported alpha-specific genes were alpha cell enriched (Fig. 3E). Several genes previously reported to be enriched in the major islet cell types, such as *MAFA*, *SLC2A2*, *SIX3*, and *DLK1* in beta cells and *IRX2*, *DPP4*, and *ADORA2A* in alpha cells, were confirmed to be signature genes. Surprisingly, we found that 37 and 33 reported beta- and alpha-specific genes were also expressed in delta and PP/gamma cells, respectively (Fig. 3F; Supplemental Table S10). Notable examples included beta and delta cell expression of the congenital hyperinsulinemia (CHI) gene *HADH* and alpha and PP/gamma cell expression of the *ARX* transcription factor (Liu et al. 2011). *HADH* is typically associated with beta cell expression and, when mutated, leads to insulin hypersecretion and CHI (Kapoor et al. 2010; Pepin et al. 2010); these data implicate the delta cell in the molecular genetics of CHI. Misexpression of *ARX* has been shown to convey both alpha and PP/gamma cell features to cells (Collombat et al. 2007), suggesting that its expression in each cell type is important for identity and function.

Genes underpinning metabolic function, regulation of energy homeostasis, and satiety are specific to distinct islet cell types

Perturbations in genes involved in glucose sensing and proper maintenance of blood glucose levels contribute to T2D pathophysiology (Schuit et al. 2001; MacDonald et al. 2005). Beta cells regulate blood glucose through the secretion of insulin and are thus exquisitely sensitive to blood glucose levels. Glucose-stimulated insulin secretion (GSIS) is linked to universal basic pathways of cellular metabolism in beta cells. To gain insight into beta cell-type–specific transcriptomic features associated with GSIS, namely, glucose uptake and glycolysis, we examined the expression of relevant genes in our islet single-cell transcriptomes (Fig. 4A).

GSIS pathway genes associated with glucose sensing and uptake displayed highly beta cell–specific expression, including *SLC2A2*, which encodes the glucose transporter GLUT2; *G6PC2*, which encodes a subunit of glucose-6-phosphatase; and *PFKFB2*, which encodes an enzyme involved in regulation of glycolysis (Fig. 4A; Chen et al. 2008; Muller et al. 2015). While expressed in all cell types, the enzyme, ALDOA1, immediately downstream from PFK1 and associated with the glycerol phosphate (GP) shuttle, is enriched in beta cells, perhaps reflecting an additional point of GSIS control. Protein-coding genes for five subsequent glycolytic enzymatic steps from glyceraldehyde-3-phosphate to pyruvate were not significantly differentially expressed between cell types. Beta cells are known to be limited in their ability to produce lactate from pyruvate (Fridlyand and Philipson 2010); this is reflected by high *LDHB*/*LDHA* ratios that favor the lactate to pyruvate flux in beta cells.

The glycerol-3-phosphate shuttle allows NAD^+ regeneration in the cytosol to sustain glycolytic flux essential for GSIS. Cytoplasmic NAD^+ generation has been shown to be essential for GSIS (Eto et al. 1999). Both components of the glycerol-3-phosphate shuttle, cytoplasmic *GPD1* and mitochondrial *GPD2*, were expressed in beta cells, with the former representing a beta cell signature gene (Fig. 4A). Additionally, we identified the mitochondrial solute transporter *SLC25A34* as beta cell specific. While its transport specificities have yet to be determined, the closest yeast ortholog of *SLC25A34*, Oac1p/YKL120w (Palmieri et al. 1999; Marobbio et al. 2008), is thought to import oxaloacetate into the mitochondria. This is particularly intriguing considering our data and others (MacDonald et al. 2011) show the complete absence of pyruvate carboxylase (PC) expression in human beta cells, despite the essential role PC is known to play in rodent GSIS (Sugden and Holness 2011) through mitochondrial production of oxaloacetate. We hypothesize that *SLC25A34* may provide an alternate, cytoplasmic source for mitochondrial oxaloacetate in the human beta cell.

Single-cell profiling also allowed us to interrogate the transcriptional repertoire of less abundant delta and PP/gamma cell types, which have been elusive in both whole islet and sorted islet studies. While it is difficult to determine epsilon cell expression signatures with one ghrelin-positive cell, our ND data set includes 16 delta cells and 12 PP/gamma cells. Among the top 100 differentially expressed (FDR < 5%) genes in delta versus other islet endocrine cells are receptors for the appetite-regulating hormones leptin (*LEPR*) and ghrelin (*GHSR*), the growth factor neuregulin 4 (*ERBB4*), and the neurotransmitter dopamine (*DRD2*) (Fig. 4B). *GHSR* has recently been shown to be specifically expressed and

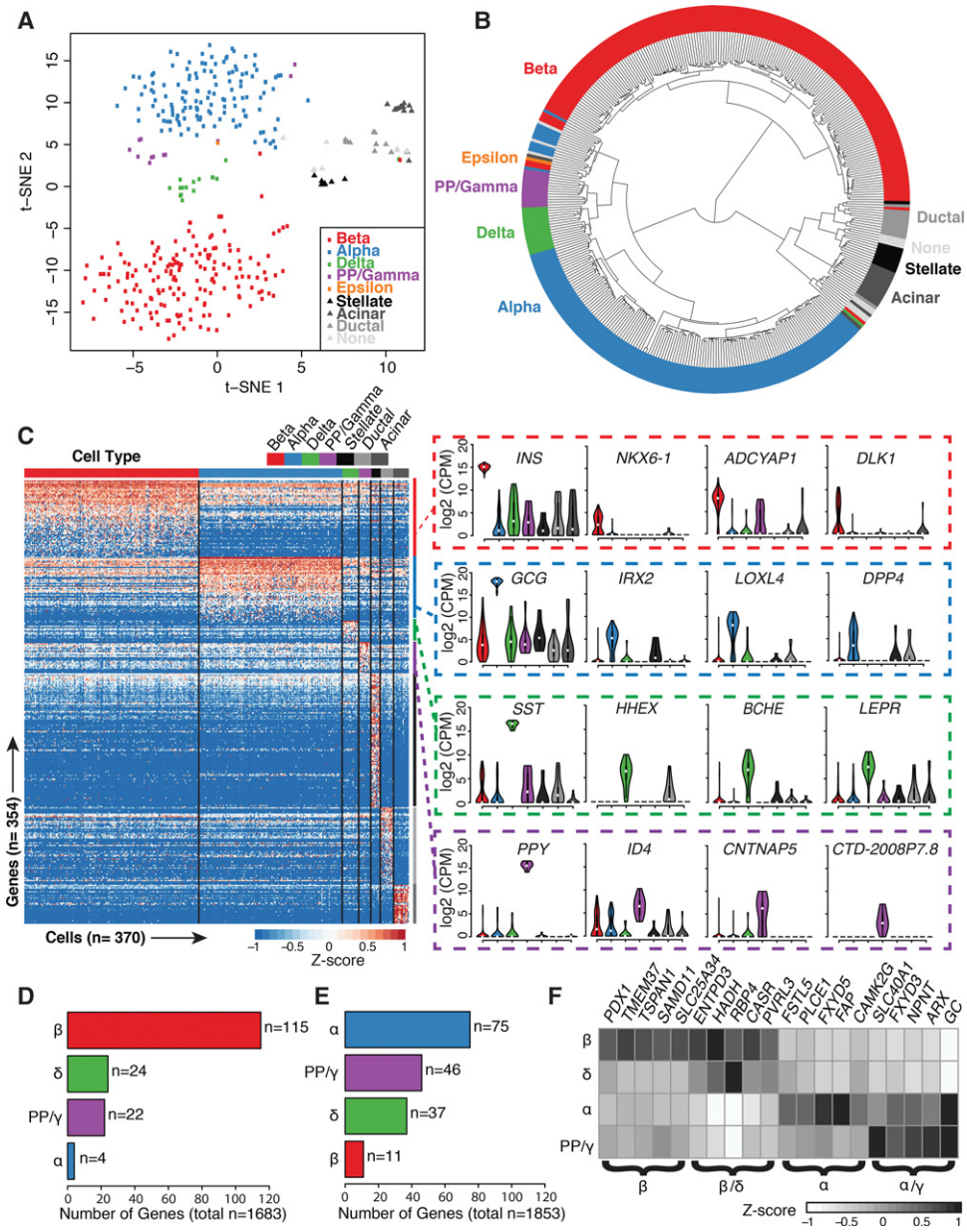


Figure 3. Statistical analysis of nondiabetic single-cell transcriptomes identifies cell-type–specific clusters and defines the signature genes of each islet cell type. (*A*) Unsupervised analysis of single-cell transcriptomes using t-distributed stochastic neighbor embedding (t-SNE) demonstrates grouping of single islet cell transcriptomes into the major constituent cell types. Respective cell labels and coloring were added after unsupervised analyses. (*B*) Unsupervised hierarchical clustering illustrates relationships of transcriptome profiles between respective endocrine and exocrine cells. (*C*) Supervised differential expression analysis of cell types determines cell-specific (signature) genes across all cells (see Methods). Values represent $\log_2$(CPM) expression after mean-centering and scaling between −1 and 1. Violin plots of selected signature gene expression are displayed to the *right* of the heatmap. (*D*,*E*) Bar plots depicting the numbers of previously reported beta-specific (*D*) and alpha-specific (*E*) genes (Dorrell et al. 2011b; Bramswig et al. 2013; Nica et al. 2013; Blodgett et al. 2015) confirmed to be expressed in each islet cell type after ANOVA and Tukey's honest significant difference (THSD) post-hoc analysis (Methods). (*F*) Several beta-specific genes demonstrate similar expression levels in delta cells, and alpha-specific genes demonstrate similar expression in PP/gamma cells. Values represent average $\log_2$(CPM) expression after mean-centering and scaling between −1 and 1. (β) Beta; (α) alpha; (δ) delta; (γ) PP/gamma cells.

functional in both human and mouse delta cells, reducing GSIS in human and mouse beta cells when induced (DiGruccio et al. 2016). *LEPR*, *DRD2*, and *ERBB4* expression is specific to human delta cells. In situ analyses (ViewRNA, Affymetrix) detected coexpression of *LEPR* in 79/102 (77%) of *SST*-expressing cells (Fig. 4D, arrowheads) in ND islets, confirming the delta cell–specific expression detected in Fluidigm C1 profiling. Thus, our data suggest intriguing roles for islet delta cells in the integration of metabolic signals via leptin, ghrelin, and dopamine signaling pathways.

PP/gamma, along with epsilon cells, are among the least studied islet cell types due to their scarcity in islets. Recent studies show that PP/gamma cells are crucial regulators of energy homeostasis (Yulyaningsih et al. 2014; Khandekar et al. 2015). In response to food intake, these cells secrete the anorexigenic hormone PPY to

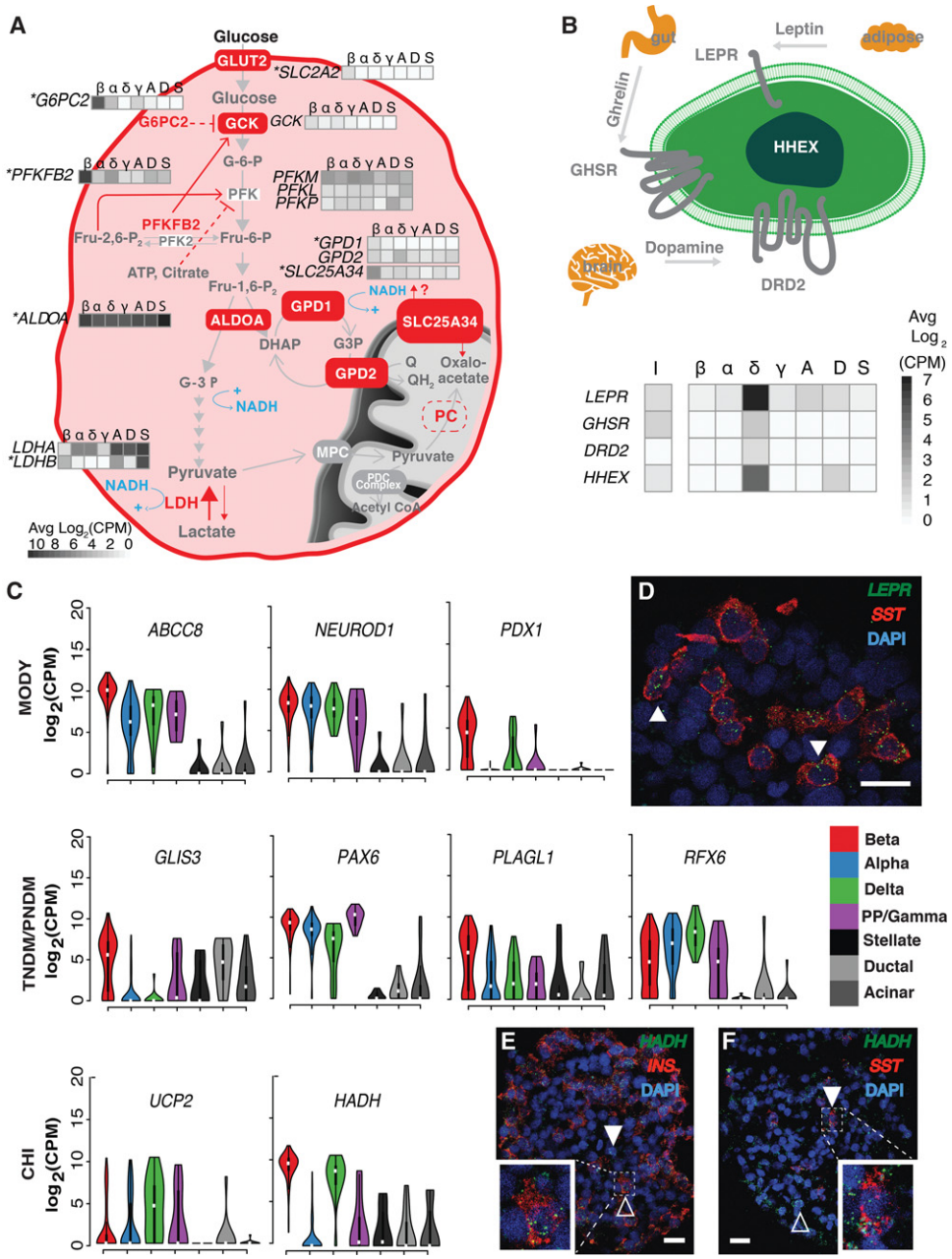


Figure 4. Cell-type–specific expression of metabolic, signaling, and diabetes trait genes. (*A*) Beta cell–specific expression of different isoforms of glycolytic and metabolic intermediate shuttles. Genes marked with an asterisk represent beta cell signature genes. (*B*) Delta cell–specific expression of neuroactive-ligand receptors and transcription factors. (I) Bulk intact islets; (β) beta; (α) alpha; (δ) delta; (γ) PP/gamma; (A) acinar; (D) ductal; (S) stellate cells. (*C*) Monogenic diabetes–associated genes and their cell-type–specific expression in islets. Violin plots show the $\log_2$(CPM) expression of each gene across cell types. (CHI) congenital hyperinsulinism; (MODY) maturity onset diabetes of the young; (TNDM) transient neonatal diabetes mellitus; (PNDM) permanent neonatal diabetes mellitus. (*D*) RNA in situ hybridization (ViewRNA, Affymetrix) of OCT-embedded islet sections from donor P3 labeling *SST* (red), *LEPR* (green), and nuclei (DAPI; blue). White arrowheads indicate $SST^+/LEPR^+$ cells. ViewRNA of OCT-embedded islet sections from donor P4 to detect the following: (*E*) *INS* (red), *HADH* (green), and nuclei (DAPI; blue) and (*F*) *SST* (red), *HADH* (green), and nuclei (DAPI; blue). White arrowheads highlight examples of $HADH^+/INS^-$ (*E*) and $HADH^+/SST^+$ (*F*) cells. Hollow arrowheads highlight $HADH^+/INS^+$ (*E*) and $HADH^+/SST^-$ (*F*) cells. In *D–F*, solid horizontal white lines indicate scale bars of 20 μm. In *E* and *F*, white dashed lines highlight a cell either co-expressing (*E*) *INS/HADH* or (*F*) *SST/HADH*. White squares in the *bottom left* of *E* and *bottom right* of *F* indicate magnified images of the cells highlighted in respective dashed white boxes.

facilitate vagal stimulation of neuropeptide Y receptors in the hypothalamus and induce satiety (Khandekar et al. 2015). Our data suggest interesting parallels in expression between PP/gamma cells and serotonergic neurons, a group of neurons that influence various cognitive and physiological processes including anxiety, mood, sleep, and satiety. We report expression of *FEV*, a serotonergic transcription factor and necessary driver of neuronal maturation previously reported in mouse beta cells (Ohta et al. 2011), in PP/gamma cells (average $\log_2$CPM of 2.172). Interestingly, FEV has also been implicated in beta cell differentiation, and *Fev*

−/− mice exhibit insulin production, insulin secretion, and glucose clearance defects (Ohta et al. 2011). Other related signature genes in PP/gamma cells include *TPH1*, encoding a tryptophan hydroxylase essential for the initial catalysis of serotonin, and *SLC6A4*, a serotonin reuptake transporter. Serotonin colocalizes with insulin in beta cells and promotes GSIS (Paulmann et al. 2009). Mice lacking *TPH1* are diabetic and exhibit impaired insulin secretion due to a lack of pancreatic serotonin (Paulmann et al. 2009). Elevated *FEV*, *TPH1*, and *SLC6A4* expression suggests PP/gamma cells share a suite of characteristic genes with serotonergic neurons that, in the pancreas, integrate central and peripheral hunger and satiety cues. We also observed high PP/gamma expression of muscarinic acetylcholine receptor M3, *CHRM3*, which stimulates exocrine pancreatic amylase (Gautam et al. 2005), insulin secretion (Kong and Tobin 2011; Molina et al. 2014), and smooth muscle contraction and gastric emptying (Eglen et al. 1994). These data implicate the less abundant delta and PP/gamma cell types as critical for islet function via the integration of systemic cues and warrant further studies to elucidate the function and health of these cells in normal and diabetogenic conditions.

Single-cell transcriptomes link rare and common diabetes genetic risk genes to islet cell types

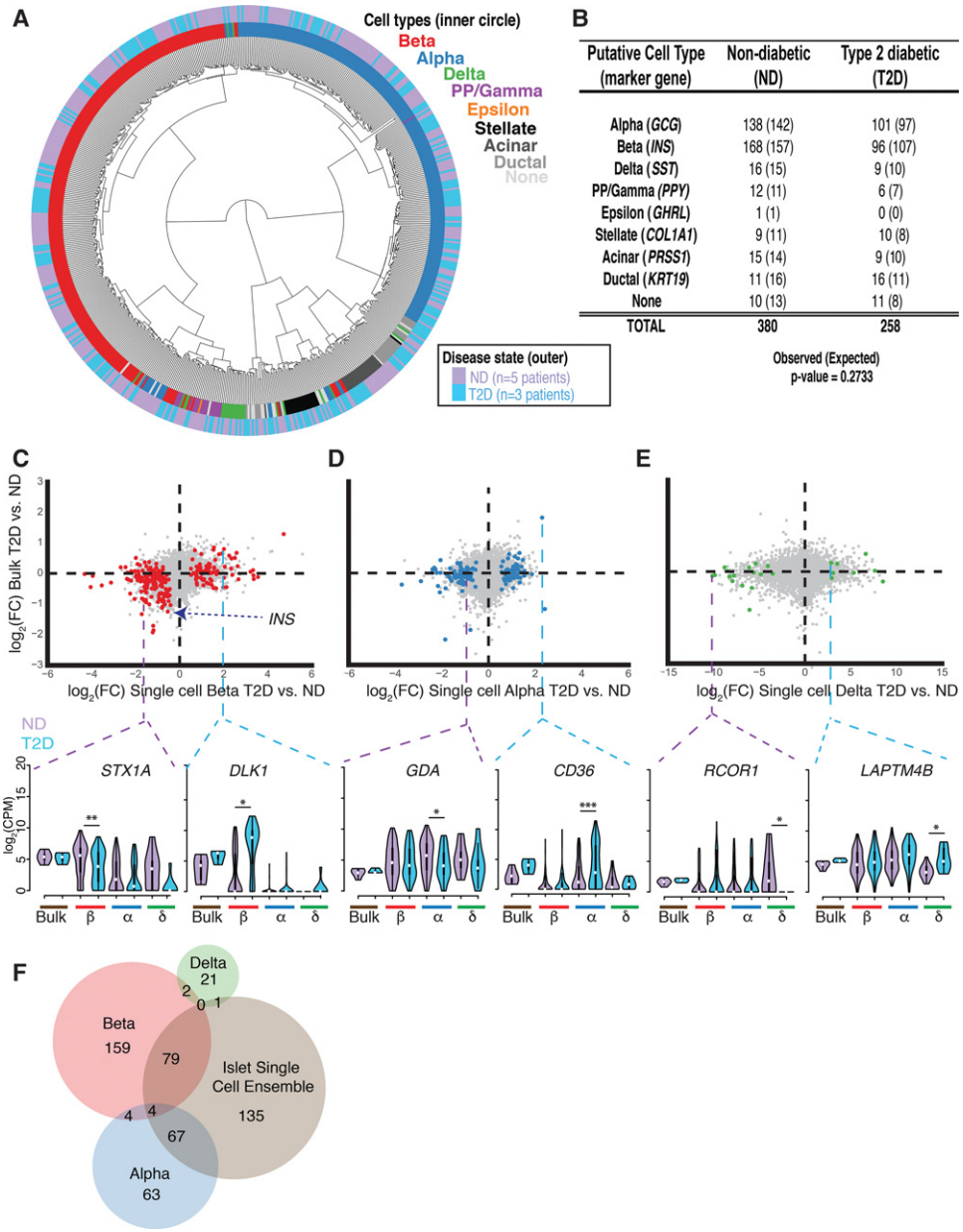
We next sought to understand the cell type(s) involved in rare forms of diabetes, including transient/permanent neonatal diabetes (T/PNDM), CHI and MODY, as well as more common forms of islet dysfunction and diabetes (T1D/T2D). Monogenic diabetic disorders, including CHI, MODY, and neonatal diabetes, are characterized by mutations in a single gene, often resulting in beta cell dysfunction and death (Schwitzgebel 2014). Five monogenic diabetes risk genes (Supplemental Table S11; Hoffmann and Spengler 2012; Senniappan et al. 2013; Schwitzgebel 2014), were enriched in beta cells (i.e., greater than fourfold change in expression in specific islet cell type relative to other endocrine cells), including glucose transporter *SLC2A2* (data not shown), beta cell maturation transcription factor *PDX1*, and the sulfonylurea drug target *ABCC8* (Fig. 4C). *PDX1* expression has been reported in human (Li et al. 2016) and mouse (DiGruccio et al. 2016) beta and delta cells. Despite the modest number of delta cells sampled, our data also suggest moderate *PDX1* expression in delta cells (four of 16 delta cells with expression ≥16 CPM). Robust expression of *HADH* in both beta and delta cells (Fig. 4C) was confirmed by in situ (View RNA) analyses (Fig. 4E,F). Approximately 386/457 cells (84%) in *HADH* and *INS* labeled sections coexpressed both markers (shown in Fig. 4E). Adjacent *SST/HADH* colabeling yielded an approximately equal proportion ($n = 255/306$; 83%) of *SST*-negative/*HADH*-positive cells. Finally, 43/457 (9%) cells were *INS* negative/*HADH* positive, and 41/306 (13%) cells coexpressed *SST* and *HADH* (shown in Fig. 4F) in the respective serial sections. Another CHI-associated gene, *UCP2* (González-Barroso et al. 2008; Senniappan et al. 2013), which was reported to be highly expressed in human beta cells (Liu et al. 2013) and to suppress insulin secretion (Krauss et al. 2003), was enriched in delta cells (Fig. 4C). Delta cell expression of monogenic diabetes genes thus implicate this cell type in the molecular genetics of rare islet dysfunction and diabetes disorders, particularly CHI.

We also investigated cell type expression patterns of 536 islet expression quantitative trait loci (eQTL) target genes (Lyssenko et al. 2009; Dupuis et al. 2010; Dayeh et al. 2013; Fadista et al. 2014; Kulzer et al. 2014; van de Bunt et al. 2015). The majority of these genes ($n = 309$; Supplemental Table S11) were lowly expressed in both the endocrine islet single-cell transcriptomes and in the paired bulk islet transcriptomes (Supplemental Fig. S16A). One hundred fifty-nine additional genes did not exhibit a greater than or equal to fourfold expression change in any endocrine islet cell type. Of the remaining 68 eQTL genes, 54, 46, 51, and 43 were expressed in beta, alpha, delta, and PP/gamma cells, respectively. Surprisingly, beta and delta cells possessed the highest numbers of cell-type–specific eQTL genes (Supplemental Table S11).

Genome-wide association studies (GWAS) have identified more than 100 loci associated with T2D and related quantitative traits (Mohlke and Boehnke 2015). Because GWAS identify genetic variants associated with a disease, but not the specific gene(s) affected (Pearson and Manolio 2008; Manolio 2010), we took two approaches to assess cell-type expression of patterns of putative GWAS genes. First, we compiled and examined a list of 197 reported putative T1D and T2D GWAS genes (Bakay et al. 2013; Nica et al. 2013; Fadista et al. 2014; Marroqui et al. 2015; Mohlke and Boehnke 2015). Of these genes, 37 were expressed in beta, 24 in alpha, 28 in delta, and 22 in PP/gamma cells (Supplemental Table S11). Similarly, genes that were cell-type specific were expressed at higher levels in ND bulk intact islets compared with those genes without cell-type specificity (Supplemental Fig. S16B). Ten genes were uniquely expressed in beta cells, including *MEG3*, a type 1 diabetes (T1D)–associated lincRNA with reported expression in mouse beta cells and potential tumor suppressor activity (Modali et al. 2015), and *IAPP*, whose protein product, when aggregated, possesses cytotoxic properties that may contribute to beta cell death and dysfunction in T2D (Westermark et al. 2011). We also identified five putative T2D GWAS genes (including *HHEX*) to be uniquely expressed in delta cells. To conduct a more liberal analysis of putative GWAS genes, we identified all single-nucleotide polymorphisms (SNPs) associated with polygenic diabetes and related traits from the GWAS catalog (https://www.ebi.ac.uk/gwas/). For each reported SNP associated with T2D, T1D, fasting insulin, fasting glucose, and proinsulin, we examined the expression of all genes overlapping within one megabase of the chromosomal locus and identified 263 genes with cell-type–specific expression (Supplemental Table S12). Together, our observations of cell-type–specific expression of eQTL and monogenic and common (T2D GWAS) diabetes genes both confirm beta cell–specific expression of multiple diabetes-associated genes (*MEG3*, *DLK1*, *SLC2A2*, etc.) and implicate other cell types in the molecular genetic pathogenesis of diabetes. In light of recent studies (Zhang et al. 2014; DiGruccio et al. 2016) and our data, which suggest that delta cells may be critical regulators of glucose homeostasis and islet function, this provides a new avenue for investigation of T2D pathogenesis, as well as potentially new therapeutic targets and treatment options.

Comparison of T2D and ND single-cell transcriptomes uncovers cell-type–specific differences not detected in whole islets

Finally, we compared single-cell transcriptome profiles from T2D and ND donors to identify differentially regulated genes and obtain greater insight into the molecular genetic pathogenesis of diabetes. After unsupervised hierarchical clustering (Fig. 5A) and t-SNE analysis (Supplemental Figs. S17, S18) using 2754 of the most highly expressed genes (Supplemental Table S4), we observed that transcriptomes clustered by cell type regardless of disease state. As previously observed, clustering was not driven by marker gene expression (Supplemental Figs. S19, S20). For regions of the dendrogram (Fig. 5A) where samples appeared to cluster by disease



Putative Cell Type (marker gene)	Non-diabetic (ND)	Type 2 diabetic (T2D)
Alpha (*GCG*)	138 (142)	101 (97)
Beta (*INS*)	168 (157)	96 (107)
Delta (*SST*)	16 (15)	9 (10)
PP/Gamma (*PPY*)	12 (11)	6 (7)
Epsilon (*GHRL*)	1 (1)	0 (0)
Stellate (*COL1A1*)	9 (11)	10 (8)
Acinar (*PRSS1*)	15 (14)	9 (10)
Ductal (*KRT19*)	11 (16)	16 (11)
None	10 (13)	11 (8)
TOTAL	380	258

Figure 5. Single-cell transcriptome analyses identify cell-type–specific expression changes in T2D islets. (*A*) T2D and ND single-cell transcriptomes cluster together by cell type after unsupervised hierarchical clustering. (*B*) Number of each ND and T2D cell type classified by marker gene expression as shown in Figure 2. The numbers of cells expected in each condition based on a χ^2 test are indicated in parentheses. (*C–E, top*) Scatter plots of $\log_2$ fold-change (FC) expression detected between T2D and ND samples from bulk intact RNA-seq (*y*-axis) and from Fluidigm C1 single-cell RNA-seq (*x*-axis) from beta cells (*left* plot; red), alpha cells (*middle* plot; blue), and delta cells (*right* plot; green). (*Bottom*) Violin plots highlight examples of differentially expressed genes in one single-cell type. Dashed purple lines represent repressed genes in the respective T2D cell type, while dashed blue lines represent induced genes. (*) FDR < 0.05, (**) FDR < 0.01, (***) FDR < 0.001. (*F*) Venn diagram showing the intersections of differentially expressed genes identified between T2D and ND transcriptomes at single-cell-type and islet single-cell ensemble resolution. The islet single-cell ensemble represents the pooled collection of beta, alpha, delta, and PP/gamma single cells.

state, we found that islet donor identity was an underlying factor that reflected sample subclustering (Supplemental Fig. S21). We obtained fewer beta cells among the T2D islet cells sampled compared with ND samples (Fig. 5B). However, observed differences in T2D and ND single-cell proportions did not differ significantly from expected cell-type proportions (Fig. 5B, χ^2 *P*-value = 0.2733), and none of the islets from these newly diagnosed T2D individuals exhibited as significant a decrease as previously reported (Butler et al. 2003; Cnop et al. 2005; Donath et al. 2005; Prentki and Nolan 2006).

Recent studies have reported features of beta cell de-differentiation under diabetogenic and stress conditions (Talchai et al. 2012; Wang et al. 2014; Cinti et al. 2016). However, we did not identify significant shifts in islet cell populations, increases in number of hormone-negative "none" cells, or appearances of new or more abundant populations of cells in T2D islets that clustered distinctly

from the known islet cell types in this study. Moreover, expression of reported de-differentiation genes including *FOXO1*, *NANOG*, and *POU5F1* (Talchai et al. 2012) did not differ significantly between T2D and ND islet cell types nor the paired bulk intact islet preparations (Supplemental Fig. S22). Finally, other de-differentiation markers such as *NEUROG3* and *MYCL* were not detected in our single-cell or bulk intact islet data. Thus, our analysis did not identify transcriptional evidence of de-differentiated cells in T2D islets.

Comparison of islet cell-type transcriptomes (e.g., T2D beta vs. ND beta) did, however, identify 410 genes that were differentially expressed (FDR < 5%) between T2D and ND donors (Supplemental Table S6) beta, (Fig. 5C, $n = 248$), alpha (Fig. 5D, $n = 138$), and delta cells (Fig. 5E, $n = 24$). We also identified differentially expressed genes in acinar ($n = 74$), ductal ($n = 35$), and stellate ($n = 28$) exocrine cell types (Supplemental Fig. S23; Supplemental Table S6). T2D beta cells exhibited a 1.4-fold decreased *INS* expression compared with ND beta cells (Fig. 5C). *STX1A* was significantly reduced ($\log_2$FC −1.5178) in T2D beta cells, consistent with reported decreases in STX1A protein levels in T2D beta cells (Andersson et al. 2012). STX1A combines with SNAP-25 and VAMP2 to form a tertiary SNARE protein complex important for insulin secretion in beta cells (Andersson et al. 2012), and *STX1A* inhibition drastically reduces GSIS and exocytosis (Vikman et al. 2006). Additionally, we detected elevated *DLK1* expression in T2D beta cells ($\log_2$FC 2.010), which has been implicated in T1D/T2D GWAS (Wallace et al. 2010) and is part of a dysregulated locus in T2D islets (Kameswaran et al. 2014). *Dlk1*$^{-/-}$ mice exhibit increased glucose sensitivity and insulin secretion (Abdallah et al. 2015), and high levels of serum DLK1 have been associated with insulin resistance in both rodents and humans (Chacón et al. 2008). Immunofluorescence indicates that DLK1 is beta cell specific in human but not mouse islets (Li et al. 2016), and FACS-enriched mouse beta cells show low expression of *Dlk1* in comparison to other sorted islet alpha and delta cells (DiGruccio et al. 2016), potentially implicating a unique role of this gene in human T2D progression. These findings suggest that perturbations in *STX1A* and *DLK1* expression may contribute to the beta cell dysfunction and impaired insulin secretion that is commonly observed in T2D pathogenesis.

Decreased beta cell function and mass are hallmarks of T2D pathophysiology (Cerf 2013; Halban et al. 2014). Our analyses suggest that transcriptional changes in nonbeta cells may also contribute to T2D pathogenesis. Specifically, we highlight increased expression of fatty acid translocase gene *CD36* ($\log_2$FC 2.296), as well as decreased expression of the guanine deaminase gene, *GDA* ($\log_2$FC −1.062), in T2D alpha cells. Soluble CD36 is a biomarker of T2D (Alkhatatbeh et al. 2013) and diabetic nephropathy (Shiju et al. 2015) and coordinates activation of the NLRP3 inflammasome, leading to proinflammatory cytokine release and reduced insulin secretion (Sheedy et al. 2013). Within T2D delta cell transcriptomes, we note increased *LAPTM4B* expression ($\log_2$FC 2.871) and drastically reduced *RCOR1* expression ($\log_2$FC −10.128). The underlying biological significance of these differentially regulated genes remains unclear and thus requires further investigation of their roles in nonbeta cell types and T2D pathology. We also compared the transcriptional differences between T2D and ND endocrine cells without first segregating them into islet cell types (334 ND and 212 T2D single-cell profiles). Approximately 66% of beta cell–specific ($n = 165/248$), 50% of alpha cell–specific ($n = 67/138$), and >90% of delta–specific ($n = 23/24$) changes in gene expression were missed when cell types were not defined and specifically compared (Fig. 5F). The decreased heterogeneity in the transcriptional profiles of cell-type–specific comparisons provides increased power to detect the transcriptomic differences and argues the importance of single-cell analysis in understanding the molecular basis of T2D.

Recent islet single-cell studies emerged while this study was under review. We therefore sought to validate our observed cell-type–specific differences in T2D islets using these independent data sets (Wang et al. 2016; Segerstolpe et al. 2016). We found that 54/77 genes and 32/171 were also significantly up- and down-regulated, respectively, in T2D beta cells in these studies ($P < 0.05$, two-sided Wilcoxon rank-sum test) (Supplemental Fig. S24A,B; Supplemental Table S13). Notably, *DLK1* consistently exhibited approximately fourfold induction in T2D beta cells in each study (Supplemental Fig. S24C,D) Similarly, 39/60 and 14/78 genes were significantly induced or repressed, respectively, in T2D alpha cells (Supplemental Fig. S24E,F). This included approximately twofold *CD36* induction in each study (Supplemental Fig. S24G,H). Validation rates for delta cells was notably lower, likely due to the relatively few cells profiled for comparison. However, we did note a significant increase ($\log_2$FC 1.203) in *LAPTM4B* in T2D delta cells from Segerstolpe et al. (2016), consistent with our data.

Discussion

In this study, we completed transcriptome profiling and analysis of 638 single islet cells from ND and T2D individuals. Single-cell RNA-seq protocols are often limited by their capture efficiency due to the fact that a limited proportion of each cell's total transcripts is represented in the final sequencing library (Liu and Trapnell 2016). Additionally, these approaches have difficulty detecting expression and changes in expression of low abundance transcripts. Despite these limitations, we observed a strong correlation between the transcriptomes of paired bulk islets and single cells, indicating these are high-quality and representative data sets. Based on single-cell transcriptome profiles, we have identified cells of each endocrine (alpha, beta, delta, PP/gamma, epsilon) and exocrine (stellate, ductal, acinar) type in the pancreas in an agnostic and data-driven manner.

This approach has defined expression signatures of each cell type with single-cell precision. Cell-type–specific expression patterns in our data such as *MAFA* in beta cells and *IRX2* in alpha cells are concordant with and extend those generated on a smaller set of cells and an independent platform (Li et al. 2016). Notably, our approach also uncovered important instances of shared expression between these cell types and the less common delta and PP/gamma islet populations, including genes mutated in CHI (*HADH*) and transcription factors regulating cell fate/identity (*ARX*). *HADH* encodes hydroxyacyl-CoA dehydrogenase, an important enzyme and negative regulator of glutamate dehydrogenase (GDH) and insulin secretion. Expression of *HADH* in islets has been shown to be beta cell specific (Kapoor et al. 2010; Pepin et al. 2010), and indeed, knockdown of *HADH* in rat 832/13 beta cells increases insulin secretion (Pepin et al. 2010). Surprisingly, our combined transcriptomic analyses and in situ (ViewRNA) validation of *HADH* revealed shared expression in beta and delta cells. These findings suggest delta cell dysfunction, in addition to beta cell dysfunction, as potential contributing factors to the development of monogenic diabetic disorders.

Most importantly, analysis of the delta and PP/gamma islet cell transcriptomes revealed cell-type–specific expression of multiple genes that suggest important roles for these cells in islet

physiology and the molecular genetics of islet dysfunction in rare (e.g., PNDM, TNDM, and MODY) and common (e.g., T2D) forms of diabetes. The novel transcriptome signatures uncovered for human delta and PP/gamma cells includes genes that strongly suggest important roles for each cell type in sensing and integrating specific systemic cues to govern islet (dys)function. This clearly warrants additional work to better understand the regulation and function of these cells in normal and diabetic states. New cell surface markers identified for each of these cell types could be used to specifically enrich and purify these populations for detailed functional analysis.

Finally, by comparing single-cell transcriptomes from T2D and ND islets, we were able to study quantitative changes in cell populations and cell-type–specific expression in T2D pathogenesis. Although not reaching statistical significance, we did observe a trend of decreased beta cells in T2D islets versus ND islets. This difference was not as pronounced as in previous reports, possibly due to the relatively modest number of cells sampled per individual. Alternatively, as most of the T2D islet single-cell transcriptomes came from newly diagnosed individuals, this difference may also reflect the shorter duration or decreased severity of T2D in these samples compared with other studies. Previous studies suggested that beta cell de-differentiation may underlie beta cell loss in T2D (Talchai et al. 2012; Wang et al. 2014; Cinti et al. 2016). However, a subsequent study comparing human islets from 14 T2D and 13 ND individuals did not identify clear evidence of this phenomenon (Butler et al. 2016). Similarly, our data do not provide transcriptome-based evidence of *trans*-differentiation or de-differentiation phenomena in T2D islets. We observed neither the appearance of new or distinct subpopulations among the T2D islet single cells nor significant changes of reported de-differentiation genes between T2D and ND cell types (e.g., T2D beta cells vs. ND beta cells), although it is possible that de-differentiated cells were simply not captured in our study. Overall, we identify 248, 138, and 24 genes exhibiting differential expression in T2D versus ND beta, alpha, and delta cells, respectively. Consistent with Simpson's paradox, approximately half of these genes in each major islet cell type (64% beta, 45% alpha) and ~90% of these in the less abundant delta cells were not detected in whole islet or single-cell islet transcriptomes when they were not stratified by cell type (Simpson 1951; Trapnell 2015). Each of these differentially regulated genes may represent important new candidate genes in T2D pathogenesis and therapeutic targeting.

Methods

Islet acquisition, processing, and dissociation

Islets were procured from ProdoLabs or the Integrated Islet Distribution Program (IIDP) and shipped in PIM(T) media (ProdoLabs) overnight on cold packs. Upon arrival, islets were washed and transferred into PIM(S) media with PIM(G) and PIM (ABS) supplements according to the manufacturer's instructions (ProdoLabs) and incubated at 37°C in a 5% CO_2 tissue culture incubator. Twenty-four hours after transfer, approximately 500 islet equivalents (IEq) were aliquoted and centrifuged at 180*g* for 3 min at room temperature (RT). One aliquot (100 IEq) was immediately flash frozen (Fig. 1A, baseline), one (200 IEq) was resuspended in 1 mL Prodo-media (Fig. 1A, intact), and one (200 IEq) was resuspended in 1 mL Accutase (Innovative Cell Technologies) (Fig. 1A, dissociated and single cell) and incubated for 10 min in a 37°C water bath, with pipetting every 2 min. Accutase-dissociated cells were filtered through a prewet cell strainer (BD) to collect single cells, rinsed with 9 mL prewarmed CMRL + 10%FBS media to stop the reaction, and centrifuged at 180*g* for 3 min at RT. Dissociated cells were resuspended in 300 µL CMRL + 10%FBS media. Cell size, number, and viability were assessed using Countess II FL (Thermo Fisher Scientific), and the cell suspension was diluted to a final concentration of 300 cells/µL. Total processing and handling time for each islet was ≤60 min.

Single-cell processing on the CI single-cell Autoprep system

After counting, cells were diluted to a final concentration range of 250–400 cells/µL and 5 µL loaded onto each C1 integrated fluidic circuit (IFC; 10- to 17-µm chip) for cell capture on the C1 single-cell Autoprep system. For each islet preparation, up to two microfluidic chips were used. After capture, cells were imaged within each capture nest with an EVOS FL auto microscope (Life Technologies). IFCs were subsequently loaded with additional reagents for subsequent cell lysis; SMARTer v1- based (Clontech), olio-(dT)-primed reverse transcription; template switching for second-strand priming; and amplification of cDNA on the C1 System. Qualitative and quantitative analysis of all single-cell cDNA products was performed on a 96 capillary fragment analyzer (Advanced Analytical). Only cell singlets, as determined by imaging, with adequate cDNA yield and quality were processed for subsequent sequencing. Fragmentation and tagmentation of cDNA was done with Nextera XT reagent (Illumina) using dual indices to prepare single-cell multiplexed libraries.

Bulk-cell RNA-seq

Bulk cells were pelleted and RNA purified using the PicoPure RNA isolation kit (Life Technologies). All RNA-seq libraries from bulk-sample RNA were generated with the same SMARTer v1 chemistry (Clontech) as for the C1 single-cell data largely following the manufacturer's instructions. Unlike the C1 workflow, after first-strand DNA synthesis, cDNA was purified using Agencourt AMPure beads (Beckman Coulter). cDNA was subsequently amplified through 12 PCR cycles. The cDNA yield and fragment size were measured on a 2100 Bioanalyzer (Agilent). For sequencing library preparation, amplified cDNA was sheared using a Covaris LE220 system to obtain fragments of ~200 bp. The fragmented cDNA was prepared for sequencing using the NEBNext DNA library prep kit for Illumina sequencing (New England Biolabs).

Sequencing, read mapping, and quality control

All sequencing was performed on a NextSeq500 (Illumina) using the 75-cycle high-output chip. RNA-seq reads were subjected to quality control using custom scripts developed at the computational sciences group at The Jackson Laboratory. Briefly, reads with >30% of bases with quality scores less than 30 were removed from the analysis, and samples with >50% of the low-quality reads were removed from the cohort. Trimmed reads were mapped to human transcriptome (GRCh37, Ensembl v70) using Bowtie 2 (Langmead and Salzberg 2012), and expression levels of all genes were estimated using RSEM (Li and Dewey 2011). Transcript per million (TPM) values as defined by RSEM were added a value of one prior to $\log_2$ transformation to avoid zeros. GRCh37 was selected for mapping to facilitate integration and comparative analyses with existing islet data sets (e.g., Parker et al. 2013; Fadista et al. 2014; van de Bunt et al. 2015) and ENCODE and NIH Roadmap data by members of the islet biology, diabetes, and functional genomics communities. The observation of expected "positive control" genes for each cell type strongly suggested that mapping to

GRCh37 instead of GRCh38 did not mask or alter any important conclusions that could be drawn from the data.

Single-cell sample processing and quality filtering

We used 26,616 protein-coding genes and lincRNAs from the GRCh37, Ensembl v70 build in our study. Genes with expression five or more TPMs in a sample were defined as expressed. Seventy-two single-cell samples that expressed fewer than 3500 genes according to these criteria were removed from downstream analysis.

Islet cell type classification

GMM of islet marker genes was performed on a per gene basis using the R-package *mclust_5.2* (Scrucca et al. 2016). Each single-cell sample was classified as a specific pancreatic cell type if and only if a single gene from the selected marker gene list—*INS* (beta), *GCG* (alpha), *SST* (delta), *PPY* (PP/gamma), *KRT19* (ductal), *PRSS1* (acinar), and *COL1A1* (stellate)—was expressed in the sample and none of the other marker genes were expressed. Cells expressing no marker genes were labeled as "none," and those expressing >1 marker gene were labeled as "multiple." Fluidigm released a white paper report detailing the potential for single cells to "*z*-stack" in up to 30% of capture nests on the medium (10–17 µm) Fluidigm C1 chip (http://info.fluidigm.com/rs/673-MRG-416/images/C1-Med-96-IFC-Redesign_wp_101-3328B1_FINAL.pdf). DAPI staining identified *z*-stacked islet cell doublets in 10% and 30% of capture nests from two additional C1 single-cell captures. Because the proportion of "multiple" labeled cells approximately equaled that of *z*-stacked doublets, we discarded these cells (n = 340) from subsequent analyses.

Unsupervised dimensionality reduction and hierarchical clustering

Barnes-Hut variant of t-SNE (van der Maaten 2014) was implemented using the R-package *Rtsne_0.10* (https://github.com/jkrijthe/Rtsne). ND single-cell transcriptomes were reduced to two dimensions in an unsupervised manner using genes with $\log_2$-CPM values greater than 10.5 in at least one sample. Similar analyses were repeated using only the T2D single-cell data and the combined single-cell data. Hierarchical clustering of cell transcriptomes was performed using Euclidean distance, Ward.D2 linkage, and the same gene selection criteria. Resultant "fan" dendrogram images were produced using the R-packages *dendextend_1.1.8* (Galili 2015) and *ape_3.5* (Paradis et al. 2004). Bulk islet transcriptomes were clustered using the same criteria.

Supervised differential gene expression analysis

Differential expression analyses were performed using the Bioconductor package *edgeR_3.14.1* (Robinson et al. 2010). Gender was used as a blocking factor to account for variability between male and female patient islets. In each comparison, protein-coding genes and lincRNAs with 20 or fewer counts in at least 20% of either cell type population being compared or at least a minimum of three cells were used. Differentially expressed genes with FDR < 5% were regarded as significant results. Endocrine cell signature genes were identified by first performing the above differential expression analysis procedure between each endocrine cell type (e.g., beta vs. alpha, beta vs. delta, and beta vs. PP/gamma). Afterward, the intersection of these results was performed to identify genes exclusively enriched in the cell type. Exocrine cell signature genes were identified after pairwise comparisons between each respective exocrine cell type (e.g., acinar vs. stellate, acinar vs. ductal). Comparisons between T2D and ND single-cell transcriptomes were performed for the same cell types (e.g., T2D beta cells vs. ND aeta cells) to identify cell-type–specific differences in gene expression between T2D and ND states.

ANOVA and post-hoc analyses

For each collection of diabetes-associated and eQTL genes examined, one-way analysis of variance (ANOVA) was used to identify statistically significant differences (FDR > 5%) in islet cell-type gene expression. Following this, we performed a post hoc analysis using a THSD test to determine genes with cell-type–specific expression patterns (fold change > 4). Genes were classified as "pan-islet" if they had an average $\log_2$(CPM) expression greater than four in all islet cell types. Genes that were not enriched in a cell type or pan-islet were classified as "lowly expressed" (average $\log_2$(CPM) < 2 in all cell types), and the remaining genes were classified as "less than fourfold change." This same methodology was used to characterize expression of the previously reported alpha- and beta-specific genes from Dorrell et al. (2011b), Bramswig et al. (2013), Nica et al. (2013), and Blodgett et al. (2015). Similar methods were used to characterize expression patterns of genes nearby diabetes-related GWAS SNPs (downloaded from the GWAS Catalog, https://www.ebi.ac.uk/gwas/, and available in Supplemental Table S12). Protein-coding RNAs and lincRNAs that overlapped within one megabase upstream of and downstream from the diabetes-associated SNPs were analyzed.

Gene set enrichment analysis

The Bioconductor package *gage_2.22.0*, (Luo et al. 2009) was used with default parameters to identify enrichment (FDR < 5%) of human Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in each of the ND islet cell transcriptomes. Enriched pathways were determined by comparing each cell-type transcriptome against the other aggregate islet cell-type transcriptomes (e.g., beta vs. alpha, delta, and PP/gamma).

RNA in situ hybridization

RNA transcripts were visualized in OCT-embedded pancreatic islet sections from two ND donors (P3 and P4) using QuantiGene ViewRNA ISH cell assay kit (catalog no. QVC0001, Affymetrix). Human *INS* ViewRNA type 6 probe (Catalog no. VA6-13248-06), *SST* ViewRNA type 6 probe (catalog no. VA6-17244-06), *LEPR* ViewRNA type 1 probe (catalog no. VA1-15221-06), and *HADH* ViewRNA type 1 probe (catalog no. VA1-12106-06) were purchased from Affymetrix (Santa Clara). The assay was performed according to the cell-based ViewRNA assay protocol with a 15-min formaldehyde fixation and a 10-min protease treatment (dilution factor 1:4000). ViewRNA probes were detected at 550 nm (Cy3) and 650 nm (Cy5) using a Leica TSC SP8 confocal microscope at 63× magnification.

Islet cell subpopulation analyses

Dorrell et al. 2016 previously defined four beta cell subpopulations with differing expression of 59 genes. With this gene set, we attempted to validate the presence of these four subpopulations via unsupervised t-SNE and hierarchical clustering of all ND beta cell transcriptomes (n = 168). Similarly, Bader et al. (2016) characterized proliferative (*Fltp*$^+$/FVR$^+$) and mature (*Fltp*$^-$/FVR$^-$) mouse beta cells that showed differential expression of 996 transcripts. By using the Mouse Genome Informatics (MGI; http://www.informatics.jax.org) database, these 996 transcripts corresponded to 691 human orthologs that were detected in our data set. Beta

cell transcriptomes were clustered using these orthologs to attempt to identify mature and proliferating subpopulations. Finally, we used the functions available in *scran_1.04* (http://bioconductor.org/packages/release/bioc/html/scran.html) to computationally assign single-cell samples into cell cycle phases (G1, G2/M, or S phase) and investigate whether our data set contained proliferating islet cells.

Data access

Raw sequence data from this study have been submitted to the databases of NCBI Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra) under accession number SRP075970 and BioProject (http://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA323853. Processed data sets from this study have been submitted to Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE86473. UCSC Genome Browser tracks of aggregate ND islet single-cell-type transcriptomes are available at http://genome.ucsc.edu/ and may be accessed with the user name "lawlorn" and session name "Islet_Single_Cell_Type_Transcriptomes." The source code used to produce the figures and tables in this paper is available in the Supplemental_Methods_Source_Code folder as suggested by Hoffman (2016).

Acknowledgments

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs, through the Peer Reviewed Medical Research Program under award no. W81XWH-16-1-0130 and by the laboratory startup funds to M.L.S. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the Department of Defense. We thank Vishal Kumar Sarsani for uploading data to the Sequence Read Archive (SRA), Jane Cha for aiding in figure design and artwork, and Anna Lisa Lucido for help with manuscript preparation. We thank Palucka laboratory members Jan Martinek and Tina Wu for help on in situ RNA analyses and members of the Stitzel, Ucar, and Parker laboratories, Travis Hinson, Karolina Palucka, and FUSION Consortium members for helpful discussions on the study.

References

Abdallah BM, Ditzel N, Laborda J, Karsenty G, Kassem M. 2015. DLK1 regulates whole-body glucose metabolism: a negative feedback regulation of the osteocalcin-insulin loop. *Diabetes* **64:** 3069–3080.

Alkhatatbeh MJ, Enjeti AK, Acharya S, Thorne RF, Lincz LF. 2013. The origin of circulating CD36 in type 2 diabetes. *Nutr Diabetes* **3:** e59.

Andersson SA, Olsson AH, Esguerra JL, Heimann E, Ladenvall C, Edlund A, Salehi A, Taneera J, Degerman E, Groop L, et al. 2012. Reduced insulin secretion correlates with decreased expression of exocytotic genes in pancreatic islets from patients with type 2 diabetes. *Mol Cell Endocrinol* **364:** 36–45.

Arda HE, Li L, Tsai J, Torre EA, Rosli Y, Peiris H, Spitale RC, Dai C, Gu X, Qu K, et al. 2016. Age-dependent pancreatic gene regulation reveals mechanisms governing human β cell function. *Cell Metab* **23:** 909–920.

Bader E, Migliorini A, Gegg M, Moruzzi N, Gerdes J, Roscioni SS, Bakhti M, Brandl E, Irmler M, Beckers J, et al. 2016. Identification of proliferative and mature β-cells in the islets of Langerhans. *Nature* **535:** 430–434.

Baetens D, Malaisse-Lagae F, Perrelet A, Orci L. 1979. Endocrine pancreas: Three-dimensional reconstruction shows two types of islets of Langerhans. *Science* **206:** 1323–1325.

Bakay M, Pandey R, Hakonarson H. 2013. Genes involved in type 1 diabetes: an update. *Genes* **4:** 499–521.

Blodgett DM, Nowosielska A, Afik S, Pechhold S, Cura AJ, Kennedy NJ, Kim S, Kucukural A, Davis RJ, Kent SC, et al. 2015. Novel observations from next-generation RNA sequencing of highly purified human adult and fetal islet cell subsets. *Diabetes* **64:** 3172–3181.

Bramswig NC, Everett LJ, Schug J, Dorrell C, Liu C, Luo Y, Streeter PR, Naji A, Grompe M, Kaestner KH. 2013. Epigenomic plasticity enables human pancreatic α to β cell reprogramming. *J Clin Invest* **123:** 1275–1284.

Brissova M, Fowler MJ, Nicholson WE, Chu A, Hirshberg B, Harlan DM, Powers AC. 2005. Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy. *J Histochem Cytochem* **53:** 1087–1097.

Butler AE, Janson J, Bonner-Weir S, Ritzel R, Rizza RA, Butler PC. 2003. β-Cell deficit and increased β-cell apoptosis in humans with type 2 diabetes. *Diabetes* **52:** 102–110.

Butler AE, Dhawan S, Hoang J, Cory M, Zeng K, Fritsch H, Meier JJ, Rizza RA, Butler PC. 2016. β-Cell deficit in obese type 2 diabetes, a minor role of β-cell dedifferentiation and degranulation. *J Clin Endocrinol Metab* **101:** 523–532.

Cabrera O, Berman DM, Kenyon NS, Ricordi C, Berggren PO, Caicedo A. 2006. The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. *Proc Natl Acad Sci* **103:** 2334–2339.

Cerf ME. 2013. β Cell dysfunction and insulin resistance. *Front Endocrinol* **4:** 37.

Chacón MR, Miranda M, Jensen CH, Fernández-Real JM, Vilarrasa N, Gutiérrez C, Näf S, Gomez JM, Vendrell J. 2008. Human serum levels of fetal antigen 1 (FA1/Dlk1) increase with obesity, are negatively associated with insulin sensitivity and modulate inflammation *in vitro*. *Int J Obes* **32:** 1122–1129.

Chen WM, Erdos MR, Jackson AU, Saxena R, Sanna S, Silver KD, Timpson NJ, Hansen T, Orrù M, Grazia Piras M, et al. 2008. Variations in the *G6PC2/ABCB11* genomic region are associated with fasting glucose levels. *J Clin Invest* **118:** 2620–2628.

Chen VP, Gao Y, Geng L, Parks RJ, Pang YP, Brimijoin S. 2015. Plasma butyrylcholinesterase regulates ghrelin to control aggression. *Proc Natl Acad Sci* **112:** 2251–2256.

Cinti F, Bouchi R, Kim-Muller JY, Ohmura Y, Sandoval PR, Masini M, Marselli L, Suleiman M, Ratner LE, Marchetti P, et al. 2016. Evidence of β-cell dedifferentiation in human type 2 diabetes. *J Clin Endocrinol Metab* **101:** 1044–1054.

Cnop M, Welsh N, Jonas JC, Jörns A, Lenzen S, Eizirik DL. 2005. Mechanisms of pancreatic β-cell death in type 1 and type 2 diabetes. *Diabetes* **54:** S97–S107.

Collombat P, Hecksher-Sørensen J, Krull J, Berger J, Riedel D, Herrera PL, Serup P, Mansouri A. 2007. Embryonic endocrine pancreas and mature β cells acquire α and PP cell phenotypes upon *Arx* misexpression. *J Clin Invest* **117:** 961–970.

Dabbs DJ. 2013. *Diagnostic immunohistochemistry*. Elsevier Health Sciences, Philadelphia, PA.

Dayeh TA, Olsson AH, Volkov P, Almgren P, Rönn T, Ling C. 2013. Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets. *Diabetologia* **56:** 1036–1046.

Dayeh T, Volkov P, Salö S, Hall E, Nilsson E, Olsson AH, Kirkpatrick CL, Wollheim CB, Eliasson L, Rönn T, et al. 2014. Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genet* **10:** e1004160.

DiGruccio MR, Mawla AM, Donaldson CJ, Noguchi GM, Vaughan J, Cowing-Zitron C, van der Meulen T, Huising MO. 2016. Comprehensive α, β and δ cell transcriptomes reveal that ghrelin selectively activates δ cells and promotes somatostatin release from pancreatic islets. *Mol Metab* **5:** 449–458.

Donath MY, Ehses JA, Maedler K, Schumann DM, Ellingsgaard H, Eppler E, Reinecke M. 2005. Mechanisms of β-cell death in type 2 diabetes. *Diabetes* **54:** S108–S113.

Dorrell C, Abraham SL, Lanxon-Cookson KM, Canaday PS, Streeter PR, Grompe M. 2008. Isolation of major pancreatic cell types and long-term culture-initiating cells using novel human surface markers. *Stem Cell Res* **1:** 183–194.

Dorrell C, Grompe MT, Pan FC, Zhong Y, Canaday PS, Shultz LD, Greiner DL, Wright CV, Streeter PR, Grompe M. 2011a. Isolation of mouse pancreatic α, β, duct and acinar populations with cell surface markers. *Mol Cell Endocrinol* **339:** 144–150.

Dorrell C, Schug J, Lin CF, Canaday PS, Fox AJ, Smirnova O, Bonnah R, Streeter PR, Stoeckert CJ Jr, Kaestner KH, et al. 2011b. Transcriptomes of the major human pancreatic cell types. *Diabetologia* **54:** 2832–2844.

Dorrell C, Schug J, Canaday PS, Russ HA, Tarlow BD, Grompe MT, Horton T, Hebrok M, Streeter PR, Kaestner KH, et al. 2016. Human islets contain four distinct subtypes of β cells. *Nat Commun* **7:** 11756.

Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, et al. 2010. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* **42:** 105–116.

Eglen RM, Reddy H, Watson N, John Challiss RA. 1994. Muscarinic acetylcholine receptor subtypes in smooth muscle. *Trends Pharmacol Sci* **15:** 114–119.

Eto K, Tsubamoto Y, Terauchi Y, Sugiyama T, Kishimoto T, Takahashi N, Yamauchi N, Kubota N, Murayama S, Aizawa T, et al. 1999. Role of NADH shuttle system in glucose-induced activation of mitochondrial metabolism and insulin secretion. *Science (New York, N.Y.)* **283:** 981–985.

Fadista J, Vikman P, Laakso EO, Mollet IG, Esguerra JL, Taneera J, Storm P, Osmark P, Ladenvall C, Prasad RB, et al. 2014. Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. *Proc Natl Acad Sci* **111:** 13924–13929.

Fridlyand LE, Philipson LH. 2010. Glucose sensing in the pancreatic β cell: a computational systems analysis. *Theor Biol Med Model* **7:** 15.

Galili T. 2015. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31:** 3718–3720.

Gautam D, Han S-J, Heard TS, Cui Y, Miller G, Bloodworth L, Wess J. 2005. Cholinergic stimulation of amylase secretion from pancreatic acinar cells studied with muscarinic acetylcholine receptor mutant mice. *J Pharmacol Exp Ther* **313:** 995–1002.

González-Barroso MM, Giurgea I, Bouillaud F, Anedda A, Bellanné-Chantelot C, Hubert L, de Keyzer Y, de Lonlay P, Ricquier D. 2008. Mutations in *UCP2* in congenital hyperinsulinism reveal a role for regulation of insulin secretion. *PLoS One* **3:** e3850.

Gunton JE, Kulkarni RN, Yim S, Okada T, Hawthorne WJ, Tseng YH, Roberson RS, Ricordi C, O'Connell PJ, Gonzalez FJ, et al. 2005. Loss of *ARNT/HIF1*β mediates altered gene expression and pancreatic-islet dysfunction in human type 2 diabetes. *Cell* **122:** 337–349.

Halban PA, Polonsky KS, Bowden DW, Hawkins MA, Ling C, Mather KJ, Powers AC, Rhodes CJ, Sussel L, Weir GC. 2014. β-Cell failure in type 2 diabetes: postulated mechanisms and prospects for prevention and treatment. *Diabetes Care* **37:** 1751–1758.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38:** 576–589.

Hoffman JI. 2016. Reproducibility: archive computer code with raw data. *Nature* **534:** 326.

Hoffmann A, Spengler D. 2012. Transient neonatal diabetes mellitus gene *Zac1* impairs insulin secretion in mice through Rasgrf1. *Mol Cell Biol* **32:** 2549–2560.

Hrvatin S, Deng F, O'Donnell CW, Gifford DK, Melton DA. 2014. MARIS: method for analyzing RNA following intracellular sorting. *PLoS One* **9:** e89459.

Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, Teichmann SA. 2016. Classification of low quality cells from single-cell RNA-seq data. *Genome Biol* **17:** 29.

Kameswaran V, Bramswig NC, McKenna LB, Penn M, Schug J, Hand NJ, Chen Y, Choi I, Vourekas A, Won KJ, et al. 2014. Epigenetic regulation of the DLK1-MEG3 microRNA cluster in human type 2 diabetic islets. *Cell Metab* **19:** 135–145.

Kapoor RR, Heslegrave A, Hussain K. 2010. Congenital hyperinsulinism due to mutations in *HNF4A* and *HADH*. *Rev Endocr Metab Disord* **11:** 185–191.

Khandekar N, Berning BA, Sainsbury A, Lin S. 2015. The role of pancreatic polypeptide in the regulation of energy homeostasis. *Mol Cell Endocrinol* **418:** 33–41.

Kong KC, Tobin AB. 2011. The role of M_3-muscarinic receptor signaling in insulin secretion. *Commun Integr Biol* **4:** 489–491.

Krauss S, Zhang C-Y, Scorrano L, Dalgaard LT, St-Pierre J, Grey ST, Lowell BB. 2003. Superoxide-mediated activation of uncoupling protein 2 causes pancreatic β cell dysfunction. *J Clin Invest* **112:** 1831–1842.

Kulzer JR, Stitzel ML, Morken MA, Huyghe JR, Fuchsberger C, Kuusisto J, Laakso M, Boehnke M, Collins FS, Mohlke KL. 2014. A common functional regulatory variant at a type 2 diabetes locus upregulates *ARAP1* expression in the pancreatic β cell. *Am J Hum Genet* **94:** 186–197.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12:** 323.

Li J, Klughammer J, Farlik M, Penz T, Spittler A, Barbieux C, Berishvili E, Bock C, Kubicek S. 2016. Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types. *EMBO Rep* **17:** 178–187.

Liu S, Trapnell C. 2016. Single-cell transcriptome sequencing: recent advances and remaining challenges. *F1000Res* **5**: 182.

Liu J, Hunter CS, Du A, Ediger B, Walp E, Murray J, Stein R, May CL. 2011. Islet-1 regulates *Arx* transcription during pancreatic islet α-cell development. *J Biol Chem* **286:** 15352–15360.

Liu J, Li J, Li W-J, Wang C-M. 2013. The role of uncoupling proteins in diabetes mellitus. *J Diabetes Res* **2013:** e585897.

Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. 2009. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10:** 161.

Lyssenko V, Nagorny CL, Erdos MR, Wierup N, Jonsson A, Spégel P, Bugliani M, Saxena R, Fex M, Pulizzi N, et al. 2009. Common variant in *MTNR1B* associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nat Genet* **41:** 82–88.

MacDonald PE, Joseph JW, Rorsman P. 2005. Glucose-sensing mechanisms in pancreatic β-cells. *Philos Trans R Soc B Lond B Biol Sci* **360:** 2211–2225.

MacDonald MJ, Longacre MJ, Stoker SW, Kendrick M, Thonpho A, Brown LJ, Hasan NM, Jitrapakdee S, Fukao T, Hanson MS, et al. 2011. Differences between human and rodent pancreatic islets: low pyruvate carboxylase, ATP citrate lyase, and pyruvate carboxylation and high glucose-stimulated acetoacetate in human pancreatic islets. *J Biol Chem* **286:** 18383–18396.

Manolio TA. 2010. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* **363:** 166–176.

Marobbio CMT, Giannuzzi G, Paradies E, Pierri CL, Palmieri F. 2008. α-Isopropylmalate, a leucine biosynthesis intermediate in yeast, is transported by the mitochondrial oxalacetate carrier. *J Biol Chem* **283:** 28445–28453.

Marroqui L, Dos Santos RS, Fløyel T, Grieco FA, Santin I, Op de Beeck A, Marselli L, Marchetti P, Pociot F, Eizirik DL. 2015. *TYK2*, a candidate gene for type 1 diabetes, modulates apoptosis and the innate immune response in human pancreatic β-cells. *Diabetes* **64:** 3808–3817.

Marselli L, Thorne J, Dahiya S, Sgroi DC, Sharma A, Bonner-Weir S, Marchetti P, Weir GC. 2010. Gene expression profiles of β-cell enriched tissue obtained by laser capture microdissection from subjects with type 2 diabetes. *PLoS One* **5:** e11499.

Mathison A, Liebl A, Bharucha J, Mukhopadhyay D, Lomberk G, Shah V, Urrutia R. 2010. Pancreatic stellate cell models for transcriptional studies of desmoplasia-associated genes. *Pancreatology* **10:** 505–516.

Modali SD, Parekh VI, Kebebew E, Agarwal SK. 2015. Epigenetic regulation of the lncRNA MEG3 and its target c-MET in pancreatic neuroendocrine tumors. *Mol Endocrinol* **29:** 224–237.

Mohlke KL, Boehnke M. 2015. Recent advances in understanding the genetic architecture of type 2 diabetes. *Hum Mol Genet* **24:** R85–R92.

Molina J, Rodriguez-Diaz R, Fachado A, Jacques-Silva MC, Berggren PO, Caicedo A. 2014. Control of insulin secretion by cholinergic signaling in the human pancreatic islet. *Diabetes* **63:** 2714–2726.

Muller YL, Piaggi P, Hanson RL, Kobes S, Bhutta S, Abdussamad M, Leak-Johnson T, Kretzler M, Huang K, Weil EJ, et al. 2015. A *cis*-eQTL in *PFKFB2* is associated with diabetic nephropathy, adiposity and insulin secretion in American Indians. *Hum Mol Genet* **24:** 2985–2996.

Nica AC, Ongen H, Irminger JC, Bosco D, Berney T, Antonarakis SE, Halban PA, Dermitzakis ET. 2013. Cell-type, allelic, and genetic signatures in the human pancreatic β cell transcriptome. *Genome Res* **23:** 1554–1562.

Nussey S, Whitehead S. 2001. *The endocrine pancreas*. BIOS Scientific Publishers, Oxford, UK.

Ohta Y, Kosaka Y, Kishimoto N, Wang J, Smith SB, Honig G, Kim H, Gasa RM, Neubauer N, Liou A, et al. 2011. Convergence of the insulin and serotonin programs in the pancreatic β-cell. *Diabetes* **60:** 3208–3216.

Palmieri L, Vozza A, Agrimi G, De Marco V, Runswick MJ, Palmieri F, Walker JE. 1999. Identification of the yeast mitochondrial transporter for oxaloacetate and sulfate. *J Biol Chem* **274:** 22184–22190.

Pandol SJ. 2011. The exocrine pancreas. In *Colloquium series on integrated systems physiology: from molecule to function to disease* (ed. Granger DN, Granger JP). Morgan & Claypool Life Sciences, San Rafael, CA.

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20:** 289–290.

Parker SC, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, van Bueren KL, Chines PS, Narisu N; NISC Comparative Sequencing Program, et al. 2013. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci* **110:** 17921–17926.

Paulmann N, Grohmann M, Voigt JP, Bert B, Vowinckel J, Bader M, Skelin M, Jevsek M, Fink H, Rupnik M, et al. 2009. Intracellular serotonin modulates insulin secretion from pancreatic β-cells by protein serotonylation. *PLoS Biol* **7:** e1000229.

Pearson TA, Manolio TA. 2008. How to interpret a genome-wide association study. *JAMA* **299:** 1335–1344.

Pepin E, Guay C, Delghingaro-Augusto V, Joly E, Madiraju SR, Prentki M. 2010. Short-chain 3-hydroxyacyl-CoA dehydrogenase is a negative regulator of insulin secretion in response to fuel and non-fuel stimuli in INS832/13 β-cells. *J Diabetes* **2:** 157–167.

Prentki M, Nolan CJ. 2006. Islet β cell failure in type 2 diabetes. *J Clin Invest* **116:** 1802–1812.

Reichert M, Rustgi AK. 2011. Pancreatic ductal cells in development, regeneration, and neoplasia. *J Clin Invest* **121:** 4572–4578.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26:** 139–140.
Schuit FC, Huypens P, Heimberg H, Pipeleers DG. 2001. Glucose sensing in pancreatic β-cells. *Diabetes* **50:** 1–11.
Schwitzgebel VM. 2014. Many faces of monogenic diabetes. *J Diabetes Investig* **5:** 121–133.
Scrucca L, Fop M, Murphy TB, Raftery AE. 2016. mclust 5: clustering, classification and density estimation using Gaussian Finite mixture models. *The R Journal* **8:** 289–317.
Segerstolpe Å, Palasantza A, Eliasson P, Andersson EM, Andréasson AC, Sun X, Picelli S, Sabirsh A, Clausen M, Bjursell MK, et al. 2016. Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell Metab* **24:** 593–607.
Senniappan S, Arya VB, Hussain K. 2013. The molecular mechanisms, diagnosis and management of congenital hyperinsulinism. *Indian J Endocrinol Metab* **17:** 19–30.
Sheedy FJ, Grebe A, Rayner KJ, Kalantari P, Ramkhelawon B, Carpenter SB, Becker CE, Ediriweera HN, Mullick AE, Golenbock DT, et al. 2013. CD36 coordinates NLRP3 inflammasome activation by facilitating the intracellular nucleation from soluble to particulate ligands in sterile inflammation. *Nat Immunol* **14:** 812–820.
Shiju TM, Mohan V, Balasubramanyam M, Viswanathan P. 2015. Soluble CD36 in plasma and urine: a plausible prognostic marker for diabetic nephropathy. *J Diabetes Complications* **29:** 400–406.
Simpson EH. 1951. The interpretation of interaction in contingency tables. *J R Stat Soc Series B (Methodol)* **13:** 238–241.
Sugden MC, Holness MJ. 2011. The pyruvate carboxylase-pyruvate dehydrogenase axis in islet pyruvate metabolism: going round in circles? *Islets* **3:** 302–319.
Talchai C, Xuan S, Lin HV, Sussel L, Accili D. 2012. Pancreatic β-cell dedifferentiation as mechanism of diabetic β-cell failure. *Cell* **150:** 1223–1234.
Taneera J, Lang S, Sharma A, Fadista J, Zhou Y, Ahlqvist E, Jonsson A, Lyssenko V, Vikman P, Hansson O, et al. 2012. A systems genetics approach identifies genes and pathways for type 2 diabetes in human islets. *Cell Metab* **16:** 122–134.
Trapnell C. 2015. Defining cell types and states with single-cell genomics. *Genome Res* **25:** 1491–1498.
van de Bunt M, Manning Fox JE, Dai X, Barrett A, Grey C, Li L, Bennett AJ, Johnson PR, Rajotte RV, Gaulton KJ, et al. 2015. Transcript expression data from human islets links regulatory signals from genome-wide association studies for type 2 diabetes and glycemic traits to their downstream effectors. *PLoS Genet* **11:** e1005694.
van der Maaten L. 2014. Accelerating T-SNE using tree-based algorithms. *J Mach Learn Res* **15:** 3221–3245.
Vikman J, Ma X, Hockerman GH, Rorsman P, Eliasson L. 2006. Antibody inhibition of synaptosomal protein of 25 kDa (SNAP-25) and syntaxin 1 reduces rapid exocytosis in insulin-secreting cells. *J Mol Endocrinol* **36:** 503–515.
Wallace C, Smyth DJ, Maisuria-Armer M, Walker NM, Todd JA, Clayton DG. 2010. The imprinted *DLK1-MEG3* gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes. *Nat Genet* **42:** 68–71.
Wang Z, York NW, Nichols CG, Remedi MS. 2014. Pancreatic β-cell dedifferentiation in diabetes and re-differentiation following insulin therapy. *Cell Metab* **19:** 872–882.
Wang YJ, Schug J, Won KJ, Liu C, Naji A, Avrahami D, Golson ML, Kaestner KH. 2016. Single cell transcriptomics of the human endocrine pancreas. *Diabetes* **65:** 3028–3038.
Westermark P, Andersson A, Westermark GT. 2011. Islet amyloid polypeptide, islet amyloid, and diabetes mellitus. *Physiol Rev* **91:** 795–826.
Xin Y, Kim J, Ni M, Wei Y, Okamoto H, Lee J, Adler C, Cavino K, Murphy AJ, Yancopoulos GD, et al. 2016. Use of the fluidigm C1 platform for RNA sequencing of single mouse pancreatic islet cells. *Proc Natl Acad Sci* **113:** 3293–3298.
Yulyaningsih E, Loh K, Lin S, Lau J, Zhang L, Shi Y, Berning BA, Enriquez R, Driessler F, Macia L, et al. 2014. Pancreatic polypeptide controls energy homeostasis via *Npy6r* signaling in the suprachiasmatic nucleus in mice. *Cell Metab* **19:** 58–72.
Zhang J, McKenna LB, Bogue CW, Kaestner KH. 2014. The diabetes gene *Hhex* maintains δ-cell differentiation and islet function. *Genes Dev* **28:** 829–834.
Zhao HL, Sui Y, Guan J, Lai FM, Gu XM, He L, Zhu X, Rowlands DK, Xu G, Tong PC, et al. 2008. Topographical associations between islet endocrine cells and duct epithelial cells in the adult human pancreas. *Clin Endocrinol* **69:** 400–406.

Received July 12, 2016; accepted in revised form November 16, 2016.

Single-cell transcriptomes identify human islet cell signatures and reveal cell-type–specific expression changes in type 2 diabetes

Nathan Lawlor, Joshy George, Mohan Bolisetty, et al.

Genome Res. published online November 18, 2016
Access the most recent version at doi:10.1101/gr.212720.116

CellView: Interactive exploration of high dimensional single cell RNA-seq data

Mohan T. Bolisetty[1], Michael L. Stitzel[1,2,3] & Paul Robson[1,2,3]

Advances in high-throughput single cell transcriptomics technologies have revolutionized the study of complex tissues. It is now possible to measure gene expression across thousands of individual cells to define cell types and states. While powerful computational and statistical frameworks are emerging to analyze these complex datasets, a gap exists between this data and a biologist's insight. The CellView web application fills this gap by providing easy and intuitive exploration of single cell transcriptome data.

Recent technological advances in single cell capture and nano-scale reactions have led to a major revolution in single cell transcriptomics[1,2,3]. Single cell datasets are analyzed using computational and statistical frameworks that enable feature (gene) selection, dimensionality reduction, clustering and differential gene expression. Multiple software packages exist that allow researchers well versed in computational analysis to perform this analysis[4–6]. However, identifying the exact parameters required for cell type identification is an iterative process greatly improved when informed by biology. In addition, interactive exploration of single cell datasets incorporating a biologist's knowledge greatly improves data interpretation, yet often such experts do not have big data handling skills.

Advances in web application frameworks and visualization methods for dense datasets facilitate the development of interactive applications to allow easy and intuitive exploration of single cell data. Here, we introduce an R Shiny[7] web application, CellView, that allows knowledge-based and hypothesis-driven exploration of processed single cell transcriptomic data. The input into CellView is an R dataset (.Rds) file

[1]The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut 06032, USA; [2]Institute for Systems Genomics, University of Connecticut, Farmington, Connecticut 06032, USA; [3]Department of Genetics & Genome Sciences, University of Connecticut, Farmington, Connecticut 06032, USA. Correspondence should be addressed to M.T.B. (mohan.bolisetty@jax.org) or P.R. (paul.robson@jax.org).

with three pre-computed data frames containing expression, clustering, and gene symbol information. This file is agnostic of upstream computational approaches providing flexibility in algorithms used to calculate these data frames. This .Rds file can be shared with the end user, eliminating the need for hosting datasets, thereby decreasing the size of a virtual machine or cloud instance required to host and use CellView. Multiple tabs allow for easy access to the data and visualization of gene expression across and within clusters, aiding cell type identification.

To illustrate the utility and power of CellView, we generated and analyzed single cell transcriptome data from peripheral blood mononuclear cells (PBMCs) using the 10X Genomics Chromium[8]. As defined by the CellRanger[8] pipeline, this data consisted of 6,554 single cells sequenced to 90.1% saturation with, on average, 824 genes and 2,077 molecules detected per cell. Dimensionality reduction using tSNE[9] was applied to genes selected by normalized dispersion, and with clustering by DBSCAN[10]. CellView automatically determines cluster numbers, updates the user interface, and renders a 3D scatter plot displaying cells clustered in tSNE space (Fig 1b) from the uploaded .rds file.

The ‘Explore’ tab provides cluster-centric exploration through three panel views. Panel 1 displays a 3D plot of a chosen gene’s expression across all cells. Panel 2 displays a 2D plot of the same gene’s expression across all cells in a single cluster, which users can select via drop-down list. Within Panel 2 users can download a .csv file of a gene-cell expression matrix by selecting cells with a square brush stroke. This provides convenient access to all genes expressed in a subset of cells. Panel 3 displays violin plots of the chosen gene’s cluster-specific expression and includes a total cell count for each cluster. *CD79A* (Fig 1c), a marker of B-cells, and *CD3D*, a marker of T-cells (Fig 1d), provide representative views of the ‘Explore’ tab.

The ‘Co-expression’ tab enables the generation of heatmaps to visualize expression of multiple genes either across all clusters, in the ‘AllClusters’ sub-menu, or on selected cells within a cluster, in the

'Selected cells' sub-menu (Fig 2a). The number of genes analyzed is only limited by legibility of the gene symbols in the resulting heatmap. This feature facilitates the use of known markers to empirically determine cell (sub)cluster identity.

The identification of doublets in single cell transcriptome data remains computationally challenging. The interactive nature of CellView aids in cell doublet identification. In the PBMC data, 'Subcluster-analysis' reveals a mixture of lymphoid and myeloid gene expression within cluster 7 suggesting this cluster consists of doublets. Cluster 7 represents 2.3% of all cells in this data set, reflecting the number of expected doublets for the quantity of cells processed in this experiment. Thus, CellView can be utilized as a tool to pre-process of single cell data and remove doublets prior to final visualization.

The 'Subcluster-analysis' tab also provides a powerful tool to identify different cell types within clusters (where trade-offs between sensitivity and specificity in the chosen clustering algorithm may be insufficient to identify unique clusters) or a continuum of states within a cell type. For example, blood monocytes span a continuum of classical, intermediate, and non-classical subtypes in flow cytometry analysis of cell surface markers CD14 and CD16[11]. Two populations of cells within a cluster can be selected by square brush strokes for differential gene expression analyses (using a likelihood test) to identify biologically informative markers. For example, monocytes occupying cluster 4 in the PBMC data appear to contain two lobes (Fig.2b). Differential expression between these two lobes using the 'Subcluster-analysis' tool identified *CD16/FCGR3A* as the most differentially expressed gene marking the smaller lobe. This lobe also contained higher expression of MHC class II genes, an additional feature of non-classical blood monocytes. *CD14* is among the top 10 up-regulated genes in the large lobe, which include other classical blood monocyte markers (e.g. *S100A8*, *S100A9*, *S100A12*). Thus, this blood monocyte continuum defined by two cell surface molecules is detected by this transcriptome cytometry approach and represented by 837 parameters (*i.e.* genes) per cell. CellView data visualization may enable immunologists to explore further underlying biology within the blood monocyte compartment, such as

investigating a subset of cells within the intermediate sub-cluster expressing *C1QA*, *C1QB*, and *C1QC*, markers of macrophage in tissue.

Dendritic cells (DCs) occupy clusters 6 and 8 in the PBMC data. Cluster 6 represents plasmacytoid DCs, expressing *CLEC4C*/*CD303*, *CD68*, *IL3RA*/*CD123* and *LILRA4*/*CD85g*. Myeloid DCs comprise cluster 8. CellView's 'Subcluster-analysis' tool enables identification of both the common CD1C+ DC (Fig. 2c; 113/215 cells expressing *CLEC10A, CD1C*) and less adundant CD141+ DCs (Fig. 2d; 12/215 cells expressing *CLEC9A, IRF8*). An additional layer of data we include in our .rds files are the genes and unique molecular identifiers (UMIs) detected per cell; this can enable identification of cell type biological features since RNA abundance (and therefore UMI count) often correlates with cell size[12]. Notably, non-classical blood monocytes and myeloid dendritic cells have the greatest numbers of UMIs detected per cell, at 3,719 and 5,645 respectively. In contrast, remaining cells have 2,305 UMIs per cell. Myeloid DCs are not noticeably larger than other PBMCs and non-classical blood monocytes are somewhat smaller in size than classical blood monocytes[11] suggesting the RNA content is reflective of an underlying biological feature of these cells rather than cell size and may reflect the precursor relationship between non-classical monocytes and myeloid dendritic cells[13].

We next applied CellView to human pancreatic islet single cell transcriptome data we generated on the Chromium system from a nondiabetic normal donor, which resulted in 4,806 cells sequenced to 87.1% saturation and detecting, on average, 1,848 genes and 7,686 molecules detected per cell. Our pipeline identified eight distinct clusters (Fig 1e). Using CellView's 'AllClusters', and marker genes we had previously used to cell type in human islets[14], we identified endocrine alpha, beta, delta, and gamma cell clusters and exocrine acinar, ductal, and stellate cell clusters. An 8th cluster represented endothelial cells (Fig 1f). Visual inspection of the 3D scatter plot displaying cells in tSNE space indicated two sub-clusters within the defined stellate cell cloud. The 'Sub-cluster' tool revealed, in addition to the stellate cells, a sub-cluster expressing the pericyte marker *RGS5*[15]. The close proximity of stellate cells and pericytes are

likely a result of their shared mesenchymal origin, as both express *COL1A1* and *ACTA2*. Visual inspection of the ductal cell cluster identified a spread of cells suggestive of a continuum of cell states. Differential expression between cells at opposing ends of this continuum using the 'Sub-cluster' tool identified biologically meaningful differences. While all cells expressed *KRT19*, there was a transition from a *REG1A*/*AMBP*-positive (Fig 1g) to a *TFF1/TFF2/TFF3/FGF19/CAECAM6*-positive (Fig 1h) population. Whether these represent different spatially localized populations of epithelial cells within the pancreatic duct or different states of activation remains to be determined, but further highlights the utility of CellView to uncover putative novel biology.

These examples illustrate how CellView provides a powerful complement to current command line approaches to cluster and identify cell types in single cell experiments. This intuitive web application enables collaboration between biologists and computational analysts and increases the value of each single cell dataset. Moreover, the CellView framework provides a useful format to present these data in an interactive manner and can be broadly applied to single cell and bulk genomics assays with count matrix and cluster information. Until a complete atlas of cell-type transcriptomes has been defined, where a reference-based approach may prove more powerful for clustering and cell type identification[16], CellView provides a useful tool to explore and characterize single cell data.

METHODS

Single cell RNA-seq - PBMCs were purchased from AllCells, thawed quickly at 37^{o}C and into DMEM supplemented with 10% FBS. Cells were quickly spun down at 400g, for 10min. Cells were washed once with 1 x PBS supplemented with 0.04% BSA and finally resuspended in 1 x PBS with 0.04% BSA. Viability was determined using trypan blue staining and measured on a Countess FL II. Briefly, 12000 cells were loaded for capture onto the Chromium System using the v2 single cell reagent kit (10X Genomics). Following capture and lysis, cDNA was synthesized and amplified (12 cycles) as per

manufacturer's protocol (10X Genomics). The amplified cDNA was used to construct an Illumina sequencing library and sequenced on a single lane of a HiSeq 4000.

Human islets from one nondiabetic deceased organ donor (UNOS ID ADIW417) were purchased from ProdoLabs and processed to obtain a single cell suspension as previously described[14]. Briefly, islets were dissociated using Accutase and filtered through a prewet cell strainer (BD) to collect single cells. The single cell suspension was prepared and loaded onto the Chromium System as described above.

FASTQ generation and Alignments - Illumina basecall files (*.bcl) were converted to fastqs using cellranger v1.3, which uses bcl2fastq v2.17.1.14. FASTQ files were then aligned to hg19 genome and transcriptome using the cellranger v1.3 pipeline, which generates a *gene vs cell* expression matrix.

Clustering and marker gene identification - Cells with less than 500 total unique transcripts were removed prior to downstream analysis. Genes for clustering were selected based on normalized dispersion analysis. Cells were clustered using Barnes Hut t-SNE[9] with the 1000 most over dispersed genes and clusters identified using DBSCAN (eps = 5.0, minpts=15). Differential gene expression was computed using edgeR[17] and signature genes defined as genes upregulated 2 fold and FDR < 0.01 in all pairwise comparisons.

Datasets and visualization – Access to CellView from: https://www.jax.org/CellOmics

ACKNOWLEDGEMENTS

We thank Karolina Palucka for helpful discussions. This work was supported by grants from NIAMS 1P50AR070594-01 (to PR), NCI 5P30CA034196-31 (PR) and The Jackson Laboratory scientific services core budget (MB, PR). Generation and analysis of the islet 10x Genomics dataset was supported by the Assistant Secretary of Defense for Health Affairs, through the Peer Reviewed Medical Research Program

under Award No. W81XWH-16-1-0130 (to MLS). Opinions, interpretations, conclusions, and recommendations are solely the responsibility of the authors and are not necessarily endorsed by the Department of Defense.

CONFLICT OF INTEREST

The authors declare no competing financial interests.

CONTRIBUTIONS

MTB and PR conceptualized the software package. MTB wrote the software package. MTB analyzed the PBMC and Pancreatic islets data. MTB, PR and MS interpreted the data to identify cell types. MTB, PR and MS wrote the manuscript.

REFERENCES

1. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161,** 1202–1214 (2015).
2. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161,** 1187–1201 (2015).
3. Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343,** 776–779 (2014).
4. Guo, M., Wang, H., Potter, S. S., Whitsett, J. A. & Xu, Y. SINCERA: A Pipeline for Single-Cell RNA-Seq Profiling Analysis. *PLoS Comput. Biol.* **11,** e1004575 (2015).
5. Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347,** 1138–1142 (2015).
6. Kiselev, V. Y. *et al.* SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* (2017). doi:10.1038/nmeth.4236
7. Chang, W., Cheng, J., Allaire, J. J., Xie, Y. & McPherson, J. shiny: Web Application Framework for

R. R package version 1.0.0. *https://CRAN.R-project.org/package=shiny* Available at: https://CRAN.R-project.org/package=shiny. (Accessed: 24th March 2017)

8. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8,** 14049 (2017).
9. Van Der Maaten, L. Accelerating t-SNE using tree-based algorithms. *J. Mach. Learn. Res.* **15,** 3221–3245 (2014).
10. Braune, C., Besecke, S. & Kruse, R. in *Partitional Clustering Algorithms* 193–213 (2014).
11. Ziegler-Heitbrock, L. *et al.* Nomenclature of monocytes and dendritic cells in blood. *Blood* **116,** e74–80 (2010).
12. Schmidt, E. E. & Schibler, U. Cell size regulation, a mechanism that controls cellular RNA accumulation: consequences on regulation of the ubiquitous transcription factors Oct1 and NF-Y and the liver-enriched transcription factor DBP. *J. Cell Biol.* **128,** 467–483 (1995).
13. Randolph, G. J., Sanchez-Schmitz, G., Liebman, R. M. & Schäkel, K. The CD16 (FcγRIII) Subset of Human Monocytes Preferentially Becomes Migratory Dendritic Cells in a Model Tissue Setting. *J. Exp. Med.* **196,** 517–527 (2002).
14. Lawlor, N. *et al.* Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. *Genome Res.* **27,** 208–222 (2017).
15. Bondjers, C. *et al.* Transcription profiling of platelet-derived growth factor-B-deficient mouse embryos identifies RGS5 as a novel marker for pericytes and vascular smooth muscle cells. *Am. J. Pathol.* **162,** 721–729 (2003).
16. Li, H. *et al.* Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat. Genet.* (2017). doi:10.1038/ng.3818
17. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).

FIGURES

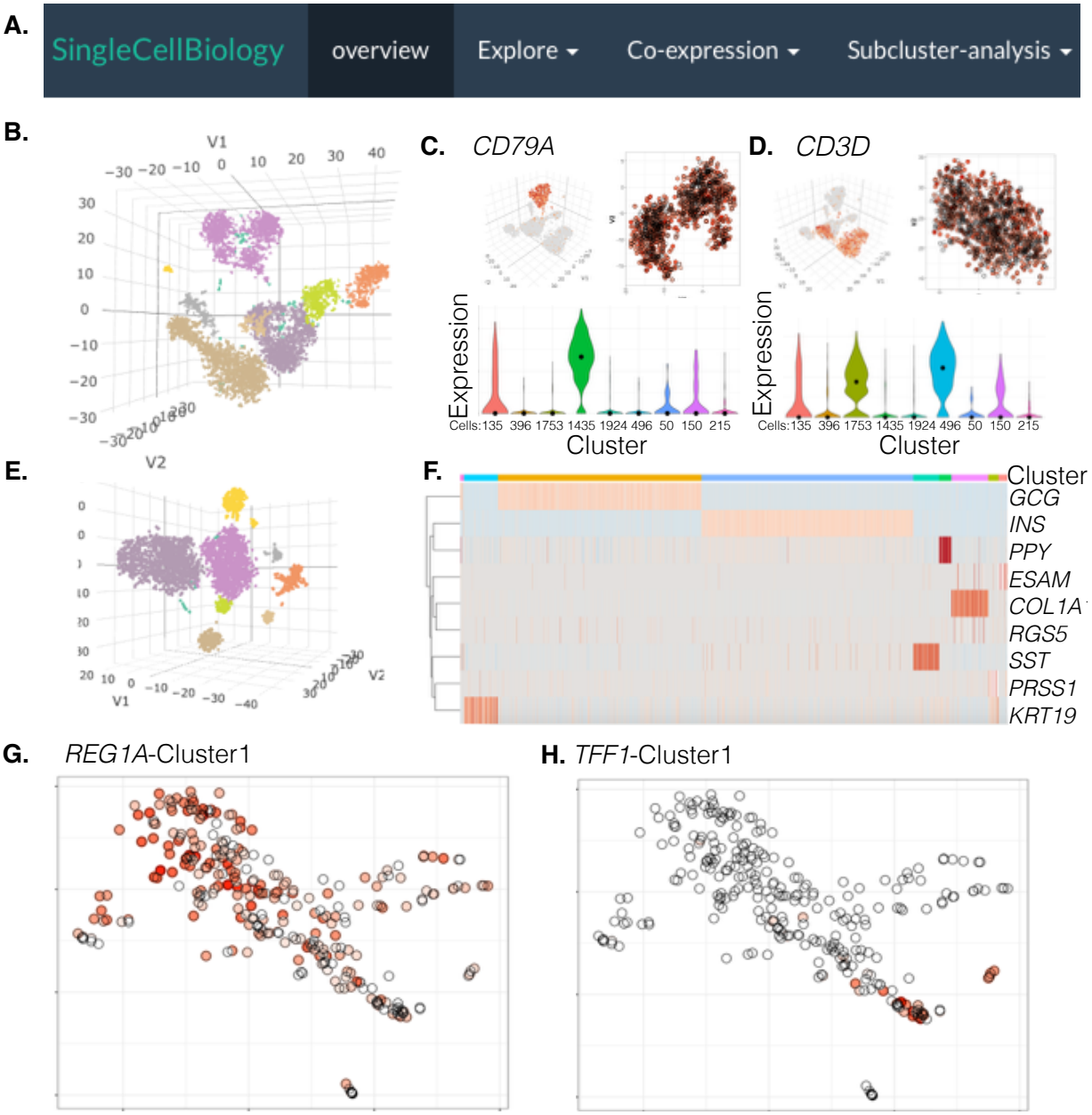


Figure 1: CellView enables cell type identification of clusters and discovery of novel cell states in PBMC and pancreatic islet datasets. **A**. CellView's graphical user interface has 3 different features that enables exploration of single cell RNA-seq datasets. **B.** Upon PBMC data upload, a 3D plot of cells clustered in t-SNE space is displayed in 'overview'. Expression patterns of marker genes such as **C**. *CD79A* and **D**. *CD3D* can be visualized in multiple panels under the 'Explore' module assisting in cell type identification and to discover further heterogeneity. **E**. 3D display of cell type clusters identified in

human pancreatic islets. **F.** Analysis using the 'Co-expression' module of CellView with marker genes aids in the identification the major endocrine cell populations, alpha (cluster 2), beta (cluster 3), gamma (cluster 5), delta (cluster 4) along with exocrine cell types like ductal (cluster 1), stellate (cluster 6), acinar (cluster 7) and endothelial (cluster 8) cells. Cluster and gene specific views, **G.** *REG1A* and **H.** *TPP1* expression in the ductal cell cluster identifies cells in multiple states.

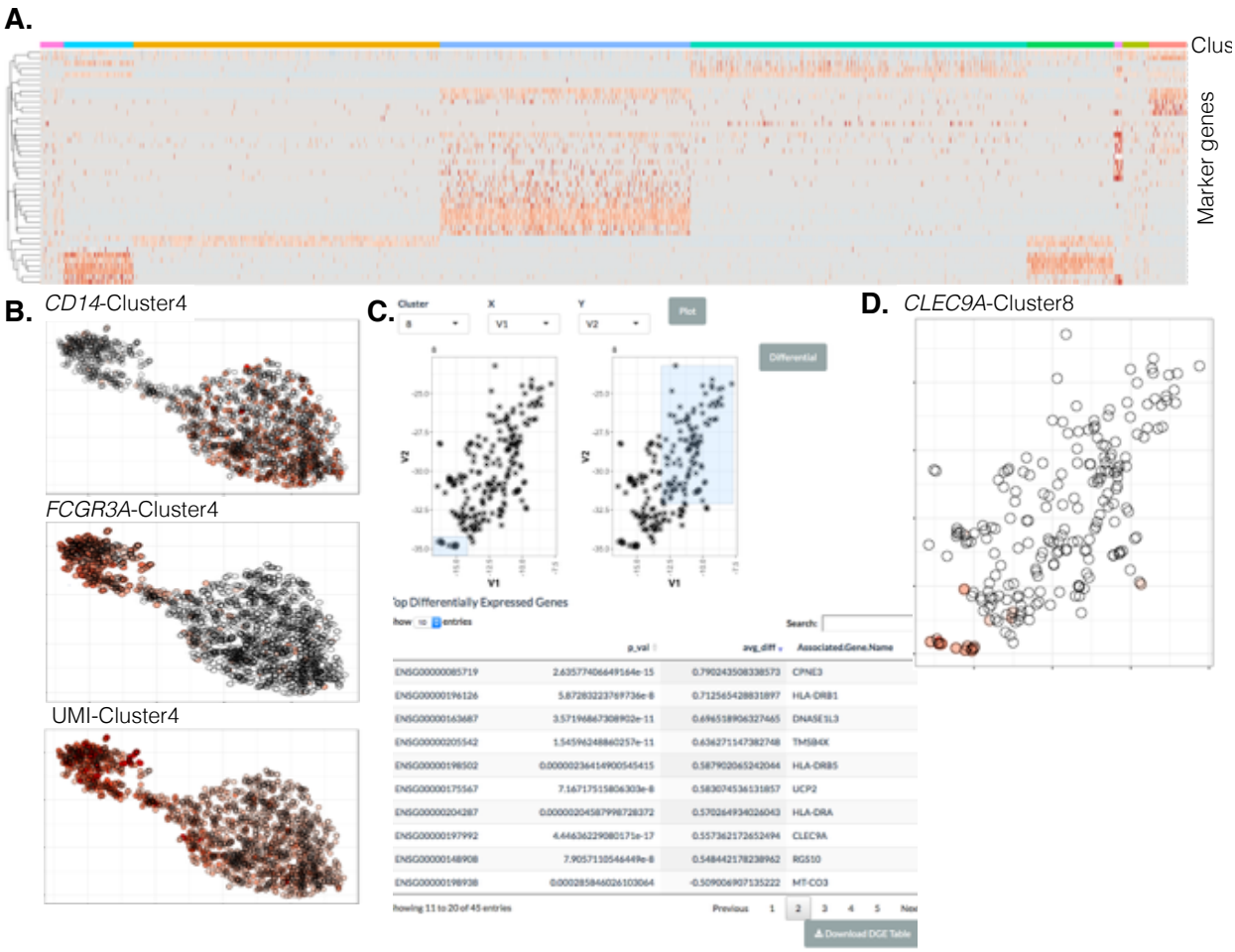


Figure 2: Investigating gene expression patterns across and within clusters with CellView identifies different cell type populations. A. Analysis using the co-expression module of CellView with various immune cell markers identifies the major subpopulations present in PBMCs. **B.** Exploring gene expression of *CD14* and *CD68* expression identifies a continuum from classical to non-classical monocytes that are also characterized by difference in absolute transcript counts (UMI). **C.** CellView

allows for identification of sub-clusters by simple differential gene expression between groups of selected cells using a square brush stroke and displays a sortable and searchable table. **D.** Differential expression of two visually resolved populations in the dendritic cell cluster identifies less abundant CD141+ dendritic cells expressing the *CLEC9A* and *DNASE1L3* marker genes.